

Isoform 1:

1 TTGCTCACTG CTCACCCACC TGCTGCTGCC ATGAGGCACC TTGGGGCCTT
 51 CCTCTTCCTT CTGGGGTCC TGGGGGCCCT CACTGAGATG TGTGAAATAC
 101 CAGAGATGGA CAGCCATCTG GTAGAGAAGT TGGGCCAGCA CCTCTTACCT
 151 TGGATGGACC GGCTTCCCT GGAGCACTTG AACCCAGCA TCTATGTGGG
 201 CCTACGCCTC TCCAGTCTGC AGGCTGGGAC CAAGGAAGAC CTCTACCTGC
 251 ACAGCCTCAA GCTTGGTTAC CAGCAGTGC TCCTAGGGTC TGCCCTCAGC
 301 GAGGATGACG GTGACTGCCA GGGCAAGCCT TCCATGGGCC AGCTGGCCCT
 351 CTACCTGCTC GCTCTCAGAG CCAACTGTGA GTTGTCAAGG GGCCACAAAGG
 401 GGGACAGGCT GGTCTCACAG CTCAAATGGT TCCTGGAGGA TGAGAAGAGA
 451 GCCATTGACA CAGCAGCCAT GGCAGGCTTG GCATTACCT GTCTGAAGCG
 501 CTCAAACTTC AACCTGGTC GGAGACAACG GATCACCAG GGCATCAGAA
 551 CAGTGGGAGA GGAGATCTTG AAGGCCCAGA CCCCCGAGGG CCACCTTG
 601 AATGTCTACA GCACCCCATT GGCATTACAG TTCCTCATGA CTTCCCCCAT
 651 GCGTGGGGCA GAACTGGGAA CAGCAGTGTCT CAAGGCGAGG GTTGCTTTGC
 701 TGGCCAGTCT GCAGGATGGA GCCTTCCAGA ATGCTCTCAT GATTTC
 751 CTGCTGCCCG TTCTGAACCA CAAGACCTAC ATTGATCTGA TCTTCCCAGA
 801 CTGCTGGCA CCACGAGTCA TGGTGGAACC AGCTGCTGAG ACCATTCTC
 851 AGACCCAAGA GATCATCAGT GTCACGCTGC AGGTGCTTAG TCTCTGC
 901 CCGTACAGAC AGTCCATCTC TGGTCTGGCC GGGTCCACCG TGGAAAGATGT
 951 CCTGAAGAAG GCCCAGGAGT TAGGAGGATT CACATATGAA ACACAGGCCT
 1001 CCTTGTCAAGG CCCCTACTTA ACCTCCGTGA TGGGGAAAGC GGCGGGAGAA
 1051 AGGGAGTTCT GGCAGCTTCT CCGAGACCCC AACACCCCAC TGTTGCAAGG
 1101 TATTGCTGAC TACAGACCCA AGGATGGAGA AACATTGAG CTGAGGCTGG
 1151 TTAGCTGGTA GCCCCCTGAGC TCCCTCATCC CAGCAGCCTC GCACACTCCC
 1201 TAGGCTTCTA CCCTCCCTCC TGATGTCCCT GGAACAGGAA CTCGC
 1251 CCTGCTGCCA CCTCCTGTGC ACTTGAGCA ATGCCCCCTG GGATCACCCCC
 1301 AGCCACAAGC CCTTCGAGGG CCCTATAACCA TGGCCACCT TGGAGCAGAG
 1351 AGCCAAGCAT CTTCCCTGGG AAGTCTTCT GGCCAAGTCT GGCCAGCCTG
 1401 GCCCTGCAGG TCTCCCATGA AGGCCACCCC ATGGTCTGAT GGGCATGAAG
 1451 CATCTCAGAC TCCCTGGCAA AAAACGGAGT CCGCAGGCCG CAGGTGTTGT
 1501 GAAGACCACT CGTTCTGTGG TTGGGGTCCT GCAAGAAGGC CTCCTCAGCC
 1551 CGGGGGCTAT GGCCCTGACC CCAGCTCTCC ACTCTGCTGT TAGAGTGGCA
 1601 GCTCCGAGCT GGTTGTGGCA CAGTAGCTGG GGAGACCTCA GCAGGGCTGC
 1651 TCAGTGCCTG CCTCTGACAA AATTAAAGCA TTGATGGCCT GTGAAAAAAA
 1701 AAAAAAAAAA AAAAAAAAAA AA

(SEQ. ID NO:1)

FEATURES:

5'UTR: 1 - 30

Start Codon: 31

Stop Codon: 1159

3'UTR: 1162

Homologous proteins:

Top 10 BLAST Hits

		Score	E
CRA 108000024653390	/altid=gi 12742775 /def=ref XP_009922.2 tr...	752	0.0
CRA 108000024636236	/altid=gi 298316 /def=gb AAB25526.1 transcr...	732	0.0
CRA 18000004926133	/altid=gi 339205 /def=gb AAA61057.1 (L02648...	732	0.0
CRA 108000024042036	/altid=gi 12654675 /def=gb AAH01176.1 AAH01...	731	0.0
CRA 18000004926130	/altid=gi 4507409 /def=ref NP_000346.1 tran...	727	0.0
CRA 18000004926132	/altid=gi 339203 /def=gb AAA61056.1 (L02647...	725	0.0
CRA 18000005170902	/altid=gi 7657639 /def=ref NP_056564.1 tran...	515	e-145
CRA 18000005218941	/altid=gi 4572454 /def=gb AAD23829.1 AF12128...	501	e-140
CRA 164000136745249	/altid=gi 11968124 /def=ref NP_071979.1 tr...	481	e-134
CRA 18000004926134	/altid=gi 4507407 /def=ref NP_001053.1 tran...	108	2e-22

FIGURE 1A

EST:

gi 10725490 /dataset=dbest /taxon=96...	858	0.0
gi 10947399 /dataset=dbest /taxon=96...	846	0.0
gi 9121897 /dataset=dbest /taxon=9606...	846	0.0
gi 13280819 /dataset=dbest /taxon=96...	846	0.0
gi 13287907 /dataset=dbest /taxon=96...	833	0.0
gi 13286505 /dataset=dbest /taxon=96...	831	0.0
gi 8150776 /dataset=dbest /taxon=960...	815	0.0
gi 5936410 /dataset=dbest /taxon=9606 ...	726	0.0
gi 6888875 /dataset=dbest /taxon=9606...	726	0.0
gi 6888872 /dataset=dbest /taxon=9606...	726	0.0

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

gi 10725490 adult adrenal gland
gi 10947399 mammary gland
gi 9121897 retinoblastoma
gi 13280819 adenocarcinoma cell line
gi 13287907 retinoblastoma
gi 13286505 embryonal carcinoma, cell line
gi 8150776 adult uterus
gi 5936410 adult uterus
gi 6888875 adult head_neck
gi 6888872 adult head_neck

Tissue Expression:

Human leukocyte

FIGURE 1B

Isoform 2:

1 GGAGGATTAA TCAGTGACAG GAAGCTGCGT CTCTCGGAGC GGTGACCAGC
 51 TGTGGTCAGG AGAGCCTCAG CAGGGCCAGC CCCAGGAGTC TTTCCCGATT
 101 CTTGCTCACT GCTCACCCAC CTGCTGCTGC CATGAGGCAC CTTGGGGCCT
 151 TCCTCTTCCT TCTGGGGTC CTGGGGGCC TCACTGAGAT GTGTGAAATA
 201 CCAGAGATGG ACAGCCATCT GGTAGAGAAG TTGGGCCAGC ACCTCTTACC
 251 TTGGATGGAC CGGCTTCCC TGAGCACTT GAACCCCAGC ATCTATGTGG
 301 GCCTACGCCT CTCCAGTCTG CAGGCTGGGA CCAAGGAAGA CCTCTACCTG
 351 CACAGCCTCA TGCTTGGTTA CCAGCAGTGC CTCCTAGGGT CTGCCTTCAG
 401 CGAGGATGAC GGTGACTGCC AGGGCAAGCC TTCCATGGGC CAGCTGGCCC
 451 TCTACCTGCT CGCTCTCAGA GCCAACTGGC ATGATCACAA GGGCCACCCC
 501 CACACTAGCT ACTACCAGTA TGCCCTGGGC ATTCTGGCCC TGTGTCTCCA
 551 CCAGAACGG GTCCATGACA CGGTGGTGA CAAACATTCTG TATGCTGTGG
 601 AACCTTCCA CCAGGGCCAC CATTCTGTGG ACACAGCAGC CATGGCAGGC
 651 TTGGCATTCA CCTGCTGAA GCGCTCAAAC TTCAACCCCTG GTCGGAGACA
 701 ACGGATCACC ATGGCCATCA GAACAGTGC AGAGGAGATC TTGAAGGCC
 751 AGACCCCCGA GGGCCACTTT GGAATGTCT ACAGCACCCC ATTGGCATTA
 801 CAGTTCTCA TGACTTCCCC CATGCGTGG GCAGAACTGG GAACAGCATG
 851 TCTCAAGGCG AGGGTTGCTT TGCTGGCCAG TCTGCAGGAT GGAGCCTTCC
 901 AGAATGCTCT CATGATTCC CAGCTGCTGC CCGTTCTGAA CCACAAGACC
 951 TACATTGATC TGATCTTCCC AGACTGTCTG GCACCCACGAG TCATGTTGGA
 1001 ACCAGCTGCT GAGACCATTG CTCAGACCCA AGAGATCATC AGTGTACCGC
 1051 TGCAGGTGCT TAGTCTTCTG CCGCCGTACA GACAGTCCAT CTCTGTTCTG
 1101 GCCGGGTCCA CCGTGGAAAGA TGTCCTGAAAG AAGGCCATG AGTTAGGAGG
 1151 ATTACATAT GAAACACAGG CCTCCTTGTC AGGCCCCCTAC TTAACCTCCG
 1201 TGATGGGAA AGCGGCCGG AAAAGGGAGT TCTGGCAGCT TCTCCGAGAC
 1251 CCCAACACCC CACTGTGCA AGGTATTGCT GACTACAGAC CCAAGGATGG
 1301 AGAAACCATT GAGCTGAGGC TGTTAGCTG GTAGCCCTG AGCTCCCTCA
 1351 TCCCAGCAGC CTCGCACACT CCCTAGGCTT CTACCCCTCCC TCCTGATGTC
 1401 CCTGGAACAG GAACTCGCCT GACCCTGCTG CCACCTCCTG TGCACCTTGA
 1451 GCAATGCCCA CTGGGATCAC CCCAGCCACA AGCCCTTCGA GGGCCCTATA
 1501 CCATGGCCCA CCTGGAGCA GAGAGCCAAG CATCTCCCT GGGAAAGTCTT
 1551 TCTGGCCAAG TCTGGCCAGC CTGGCCCTGC AGGTCTCCCA TGAAGGCCAC
 1601 CCCATGGTCT GATGGGCATG AAGCATCTCA GACTCCTTGG CAAAAAACGG
 1651 AGTCCGCAGG CCGCAGGTGT TGTGAAGACCC ACTCGTTCTG TGGTTGGGGT
 1701 CCTGCAAGAA GGCCTCTCA GCCCGGGGGC TATGGCCCTG ACCCCAGCTC
 1751 TCCACTCTGC TGTTAGAGTG GCAGCTCCGA GCTGGTTGTG GCACAGTAGC
 1801 TGGGGAGACC TCAGCAGGGC TGCTCAGTGC CTGCCTCTGA CAAAATTAAA
 1851 GCATTGATGG CCTGTAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAA

(SEQ ID NO:2)

FEATURES:

5'UTR: 1 - 131
 Start Codon: 132
 Stop Codon: 1332
 3'UTR: 1335

Homologous proteins:

Top 10 BLAST Hits

		Score	E
CRA 108000024636236	/altid=gi 298316 /def=gb AAB25526.1 transcr...	793	0.0
CRA 108000024653390	/altid=gi 12742775 /def=ref XP_009922.2 transl...	793	0.0
CRA 18000004926133	/altid=gi 339205 /def=gb AAA61057.1 (L02648...	792	0.0
CRA 108000024042036	/altid=gi 12654675 /def=gb AAH01176.1 AAH01...	792	0.0
CRA 18000004926130	/altid=gi 4507409 /def=ref NP_000346.1 transl...	788	0.0
CRA 18000004926132	/altid=gi 339203 /def=gb AAA61056.1 (L02647...	786	0.0
CRA 18000005170902	/altid=gi 7657639 /def=ref NP_056564.1 transl...	561	e-159

FIGURE 1C

CRA 164000136745249 /altid=gi 11968124 /def=ref NP_071979.1 tr...	554	e-156
CRA 18000005218941 /altid=gi 4572454 /def=gb AAD23829.1 AF12128...	545	e-154
CRA 18000004926134 /altid=gi 4507407 /def=ref NP_001053.1 tran...	128	1e-28

EST:

gi 10725490 /dataset=dbest /taxon=96...	858	0.0
gi 5936410 /dataset=dbest /taxon=9606 ...	835	0.0
gi 6888875 /dataset=dbest /taxon=9606...	726	0.0
gi 6888872 /dataset=dbest /taxon=9606...	726	0.0
gi 12258937 /dataset=dbest /taxon=960...	686	0.0
gi 10947399 /dataset=dbest /taxon=96...	680	0.0
gi 13287907 /dataset=dbest /taxon=96...	680	0.0
gi 9121897 /dataset=dbest /taxon=9606...	680	0.0
gi 13280819 /dataset=dbest /taxon=96...	680	0.0
gi 8150776 /dataset=dbest /taxon=960...	656	0.0

EXPRESSION INFORMATION FOR MODULATORY USE:

library source:

gi 10725490 adult adrenal gland
gi 5936410 adult uterus
gi 6888875 adult head_neck
gi 6888872 adult head_neck
gi 12258937 adult lung_tumor
gi 10947399 mammary gland
gi 13287907 retinoblastoma
gi 9121897 retinoblastoma
gi 13280819 adenocarcinoma cell line
gi 8150776

Tissue Expression:

Human hippocampus

FIGURE 1D

Isoform 1:

1 MRHLGAFLL LGVLGALTEM CEIPEMDSHL VEKLGQHLLP WMDRLSLEHL
51 NPSIYVGLRL SSLQAGTKED LYLHSLKLGY QQCLLGSAFS EDDGDCQGKP
101 SMGQLALYLL ALRANCEFVR GHKGDRLLVSQ LKWFLEDEKR AIDTAAMAGL
151 AFTCLKRSNF NPGRRQRITM AIRTVREEIL KAQTPEGHFG NVYSTPLALQ
201 FLMTSPMRGA ELGTACLKAR VALLASLQDG AFQNALMISQ LLPVLNHKTY
251 IDLIFPDCLA PRVMLEPAAE TIPQTQEIIIS VTLQVLSLLP PYRQSISVLA
301 GSTVEDVLKK AHELGGFTYE TQASLSGPYL TSVMGKAAGE REFWQLLRDP
351 NTPLLQGIAD YRPKDGETIE LRLVSW

(SEQ ID NO:3)

FEATURES:

Functional domains and key regions:

PDOC00005 PS00005 PKC_PHOSPHO_SITE

Protein kinase C phosphorylation site

Number of matches: 2

1	75-77	SLK
2	174-176	TVR

PDOC00006 PS00006 CK2_PHOSPHO_SITE

Casein kinase II phosphorylation site

Number of matches: 6

1	67-70	TKED
2	90-93	SEDD
3	174-177	TVRE
4	226-229	SLQD
5	249-252	TYID
6	302-305	STVE

PDOC00008 PS00008 MYRISTYL

N-myristoylation site

Number of matches: 7

1	12-17	GVLGAL
2	57-62	GLRLSS
3	86-91	GSAFSE
4	149-154	GLAFTC
5	190-195	GNVYST
6	209-214	GAELGT
7	230-235	GAFQNA

PDOC00009 PS00009 AMIDATION

Amidation site

162-165 PGRR

SignalP results:

Measure	Position	Value	Cutoff	Conclusion
max. C	19	0.602	0.37	YES
max. Y	19	0.702	0.34	YES
max. S	5	0.974	0.88	YES
mean S	1-18	0.949	0.48	YES

Most likely cleavage site between pos. 18 and 19: ALT-EM

BLAST Alignment to Top Hit:

```
>CRA|108000024636236 /altid=gi|298316 /def=gb|AAB25526.1|
  transcobalamin II, TC II [human, endothelial cells,
  Peptide, 427 aa] /org=human /taxon=9606 /dataset=nraa
  /length=427
  Length = 427
```

Score = 732 bits (1870), Expect = 0.0

Identities = 376/427 (88%), Positives = 376/427 (88%), Gaps = 51/427 (11%)

Frame = +1

FIGURE 2A

Query: 31 MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPWMDRLSLEHLNPSIYVGLRL 210
 MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPWMDRLSLEHLNPSIYVGLRL
 Sbjct: 1 MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPWMDRLSLEHLNPSIYVGLRL 60
 Sbjct: 1 MRHLGAFLFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPWMDRLSLEHLNPSIYVGLRL
 Query: 211 SSLQAGTKEDLYLHSLKLGYQQCLLGSASFSEDDGDCQGKPSMGQLALYLLALRANCEFVR 390
 SSLQAGTKEDLYLHSLKLGYQQCLLGSASFSEDDGDCQGKPSMGQLALYLLALRANCEFVR
 Sbjct: 61 SSLQAGTKEDLYLHSLKLGYQQCLLGSASFSEDDGDCQGKPSMGQLALYLLALRANCEFVR 120
 Sbjct: 61 SSLQAGTKEDLYLHSLKLGYQQCLLGSASFSEDDGDCQGKPSMGQLALYLLALRANCEFVR
 Query: 391 GHKGDRQLVSQLKWFLEDEKRAI----- 456
 GHKGDRQLVSQLKWFLEDEKRAI
 Sbjct: 121 GHKGDRQLVSQLKWFLEDEKRAIGHDHKGHPHTSYYQYGLGILALCLHQKRVHDSVVDKLL 180
 Sbjct: 121 GHKGDRQLVSQLKWFLEDEKRAIGHDHKGHPHTSYYQYGLGILALCLHQKRVHDSVVDKLL
 Query: 457 -----DTAAMAGLAFTCLKRSNFNPGRQRITMAIRTVREEILKAQTPEGHF 597
 DTAAMAGLAFTCLKRSNFNPGRQRITMAIRTVREEILKAQTPEGHF
 Sbjct: 181 YAVEPFHQGHHSVDTAAMAGLAFTCLKRSNFNPGRQRITMAIRTVREEILKAQTPEGHF 240
 Sbjct: 181 YAVEPFHQGHHSVDTAAMAGLAFTCLKRSNFNPGRQRITMAIRTVREEILKAQTPEGHF
 Query: 598 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT 777
 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT
 Sbjct: 241 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT 300
 Sbjct: 241 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT
 Query: 778 YIDLIFPDCLAPRVMLEPAAETIPQTQEIIISVTLQVLSLLPPYRQSIISVLAGSTVEDVLK 957
 YIDLIFPDCLAPRVMLEPAAETIPQTQEIIISVTLQVLSLLPPYRQSIISVLAGSTVEDVLK
 Sbjct: 301 YIDLIFPDCLAPRVMLEPAAETIPQTQEIIISVTLQVLSLLPPYRQSIISVLAGSTVEDVLK 360
 Sbjct: 301 YIDLIFPDCLAPRVMLEPAAETIPQTQEIIISVTLQVLSLLPPYRQSIISVLAGSTVEDVLK
 Query: 958 KAHELGGFTYETQASLSGPYLTSMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI 1137
 KAHELGGFTYETQASLSGPYLTSMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI
 Sbjct: 361 KAHELGGFTYETQASLSGPYLTSMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI 420
 Sbjct: 361 KAHELGGFTYETQASLSGPYLTSMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI
 Query: 1138 ELRLVSW 1158
 ELRLVSW
 Sbjct: 421 ELRLVSW 427
 (SEQ ID NO:6)

HMM results:

Model	Description	Score	E-value	N
PF01122	Eukaryotic cobalamin-binding protein	829.9	8.6e-246	2
CE00052	CE00052 lymphocyte_transmembrane_protein_KAP	3.2	2.9	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00052	1/1	1	11	[.	1	11	[. 3.2 2.9
PF01122	1/2	1	142	[.	1	143	[. 296.0 4.6e-85
PF01122	2/2	143	376	.]	197	450	.] 531.8 4.8e-156

FIGURE 2B

Isoform 2:

1 MRHLGAFLFL LGVLGALTEM CEIPEMDSHL VEKLGQHLLP WMDRLSLEHL
51 NPSIYVGLRL SSLQAGTKED LYLHSIMLGY QQCLLGSafs EDDGDCQGKP
101 SMGQLALYLL ALRANWHDHK GHPHTSYQQY GLGILALCLH QKRVHDSVVD
151 KLLYAVEPFH QGHHSVDTAA MAGLAFTCLK RSNFNPGRQQ RITMAIRTVR
201 EEEILKAQTPE GHFGNVYSTP LALQFLMTSP MRAEELGTAC LKARVALLAS
251 LQDGAFQNAL MISQLLPVLN HKTYIDLIFP DCLAPRVMLE PAAETIPQTO
301 EIISVTLQVL SLLPPYRQSI SVLAGSTVED VLKKAHELGG FTYETQASLS
351 GPYLTTSVMGK AAGEREFWQL LRDPNTPLLQ GIADYRPKDGE ETIELRLVSW
(SEQ ID NO:4)

FEATURES:

Functional domains and key regions:

PDOC00005 PS00005 PKC_PHOSPHO_SITE
Protein kinase C phosphorylation site
198-200 TVR

PDOC00006 PS00006 CK2_PHOSPHO_SITE
Casein kinase II phosphorylation site

Number of matches: 7

1	67-70	TKED
2	90-93	SEDD
3	147-150	SVVD
4	198-201	TVRE
5	250-253	SLQD
6	273-276	TYID
7	326-329	STVE

PDOC00008 PS00008 MYRISTYL

N-myristoylation site

Number of matches: 7

1	12-17	GVLGAL
2	57-62	GLRLSS
3	86-91	GSAFSE
4	173-178	GLAFTC
5	214-219	GNVYST
6	233-238	GAELGT
7	254-259	GAFQNA

PDOC00009 PS00009 AMIDATION

Amidation site

186-189 PGRR

PDOC00428 PS00468 COBALAMIN_BINDING

Eukaryotic cobalamin-binding proteins signature

165-178 SVDTAAMAGLAFTC

SignalP results:

Measure	Position	Value	Cutoff	Conclusion
max. C	19	0.602	0.37	YES
max. Y	19	0.702	0.34	YES
max. S	5	0.974	0.88	YES
mean S	1-18	0.949	0.48	YES

Most likely cleavage site between pos. 18 and 19: ALT-EM

BLAST Alignment to Top Hit:

>CRA|108000024636236 /altid=gi|298316 /def=gb|AAB25526.1|
transcobalamin II, TC II [human, endothelial cells,
Peptide, 427 aa] /org=human /taxon=9606 /dataset=nraa
/length=427
Length = 427

FIGURE 2C

Score = 793 bits (2026), Expect = 0.0
 Identities = 399/427 (93%), Positives = 399/427 (93%), Gaps = 27/427 (6%)

Query: 1 MRHLGAFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPWMDRLSLEHLPNSIYVGLRL 60
 MRHLGAFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPWMDRLSLEHLPNSIYVGLRL
 Sbjct: 1 MRHLGAFLLGVLGALTEMCEIPEMDSHLVEKLGQHLLPWMDRLSLEHLPNSIYVGLRL 60

Query: 61 SSLQAGTKEDELYLHSLMLGYQQCQLLSAFSEDDGDCQGKPSMGQLALYLLALRAN----- 115
 SSLQAGTKEDELYLHSL LGYQQCQLLSAFSEDDGDCQGKPSMGQLALYLLALRAN
 Sbjct: 61 SSLQAGTKEDELYLHSLKLGYQQCQLLSAFSEDDGDCQGKPSMGQLALYLLALRANCEFVR 120

Query: 116 -----W-----HDHKGHPHTSYYQYGLGILALCLHQKRVHDSVVDKLL 153
 W HDHKGHPHTSYYQYGLGILALCLHQKRVHDSVVDKLL
 Sbjct: 121 GHKGDRQLSQLKWFLEDEKRAIGHDHKGPHTSYYQYGLGILALCLHQKRVHDSVVDKLL 180

Query: 154 YAVEPFHQGHHSVDTAAMAGLAFTCLKRSNFPGRQRITMAIRTREEILKAQTPEGHF 213
 YAVEPFHQGHHSVDTAAMAGLAFTCLKRSNFPGRQRITMAIRTREEILKAQTPEGHF
 Sbjct: 181 YAVEPFHQGHHSVDTAAMAGLAFTCLKRSNFPGRQRITMAIRTREEILKAQTPEGHF 240

Query: 214 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT 273
 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT
 Sbjct: 241 GNVYSTPLALQFLMTSPMRGAELGTACLKARVALLASLQDGAFQNALMISQLLPVLNHKT 300

Query: 274 YIDLIFPDCLAPRVMLEPAAAETIPQTQEIIISVTLQVLSLLPPYRQSISVLAGSTVEDVLK 333
 YIDLIFPDCLAPRVMLEPAAAETIPQTQEIIISVTLQVLSLLPPYRQSISVLAGSTVEDVLK
 Sbjct: 301 YIDLIFPDCLAPRVMLEPAAAETIPQTQEIIISVTLQVLSLLPPYRQSISVLAGSTVEDVLK 360

Query: 334 KAHELGGFTYETQASLSGPYLTSMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI 393
 KAHELGGFTYETQASLSGPYLTSMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI
 Sbjct: 361 KAHELGGFTYETQASLSGPYLTSMGKAAGEREFWQLLRDPNTPLLQGIADYRPKDGETI 420

Query: 394 ELRLVSW 400
 ELRLVSW
 Sbjct: 421 ELRLVSW 427
 (SEQ ID NO:7)

HMM results:

Model	Description	Score	E-value	N
PF01122	Eukaryotic cobalamin-binding protein	906.3	8.6e-269	2
CE00052	CE00052 lymphocyte_transmembrane_protein_KAP	3.2	2.9	1

Parsed for domains:

Model	Domain	seq-f	seq-t	hmm-f	hmm-t	score	E-value
CE00052	1/1	1	11	[.	1	11	[. 3.2 2.9
PF01122	1/2	1	115	[.	1	115	[. 241.3 1.4e-68
PF01122	2/2	117	400	.]	145	450	.] 660.5 8.7e-195

FIGURE 2D

1 ATATGTATGG GAAATATGCT GTCTTCCTAT TCCTACTCCC CCACCCCTCTA
 51 GCACTGAGTC CAGGTAGTA GCCAGGGGG TGTCTCCCTC CTTTACTTCG
 101 ACACCCCTAAC TACCTTGGGG ATCAGAAGTG ACTCTCTGGA AGGATGCTGC
 151 TGCTTCTCAC CAGAGGCTGA CGATAACGAA GGCTATCCTC CATGGCCACC
 201 TCCTCCAGGC TGCCCTCCTG GAAATAGGAA TCATAATAGT TGTTACTGGA
 251 AACAGGCAGA GGGTTGGGG AGCCAAGGCA GTCCCACCA GGACCAAGGT
 301 GGCTCATTG CACACACTTC ACCATGACTC CCCTGAAGGT CCAAACGTGC
 351 GGTTCTCGGG AAGTTGGCT CCCCACTGGC CTCCCTCCTT CCTCAGAACCC
 401 TCCAGGGGTG CTCCCTCTAG TGGCCACATC CAGCCTTCT GACTGGACAA
 451 CCTATCATTT AAAATTTCAGTAGTCTGG TAAACAGACA CACGTTGCTG
 501 TATTATTTA TGTCAGGGC TTGGTTTGTG ATAAGTCAGG CTCAAAAAGA
 551 TTGTCTTAAAGAGTGAACC TTGGCAATT ACCATAAAAT AATTGCAATG
 601 CAGATTGTGC ATGAAATGA TTGGAGATAT TTTAAGGTCA TAGTGTCTTC
 651 ACAAAATTGAG CTGAAAGGG ACTGTTAGGA TGATCTTGCC TAACCCCTCTC
 701 ATCTCACACAGAAGAAC TTTTAAACTC GAGAGGTAA GTGACCTGGC
 751 CAAAGTCACA CAGGCCAAC TAGTTAACCTC GTATACATTG ATTCTCCCTGT
 801 GGGGCTGGC AGATGAGGA TCTTTTGTTC TCTTCCCTGT TTGCAGAGAT
 851 TTTTTTGAG GTTACTTTCC GAGTTCTGGC AAGTACCCCT GCTTCTGGTA
 901 GCTTTGTGTC TCGATTCAAT CTCATTCTTT TTATTTTATT TTATTTTGA
 951 GACAGGGTCT CACTTGTCA CCCAAGCTGG AGTGCAGTGG TGTAATCTTG
 1001 GCTCACTGTA GCCTCCACCT CTTGGGTTCA AGCGATCCTC CTGCCTCAGC
 1051 CCCCCAAAGTA GCTGGGATTA CAGACGCTGC CCACCACGCC AGGCTAATTT
 1101 ATGGTTTTTGTATGTGTTT TTGTGTTTT TGAGAGACA GTGTTTCCCC
 1151 ATGTTGCCCA GGCTGGTCTC CAACTCTGTA GCTCAAGTGA TCTGCCCGCC
 1201 TCAGCCTTCAAAAGTGTAG GATTACAGGT GTGAGGCCAC GTGCCCGGAC
 1251 TTAATCCCAT TCTTTAACCTT GTTTGTTTT GTCCCTCTCCA GGAGGCTCCC
 1301 AGCCCTTCG GATTGGTTGA GAAAAGTGGC CTGGCTGGTC TGGGGCCAGC
 1351 AGCACCCACC CTCCCTCAA TTGCCCAACT CCCCCCCCCA CCGAACGCC
 1401 CAACTCCCCC TCCCAACTG CCCAACTCCC CCACCCCCAC AATCCCCCTCC
 1451 CGGCCAACACT GAGGGAGGGC GTGCTGAAAA ACAGCTGACT CCAGCAATGC
 1501 TGCTCACGT ACCACTGCG CTGCAGCTCC CGTTCCACTC CTTGTCCTGG
 1551 GCTAGGTGGG CACTACCAGG GGCTCCTTGTG GTAAGGAGTA CCGGGTAGGC
 1601 ACCCGGTCTT GCCAATCCAC CACTGGAAACA GCTGGGGGAG CAGCAGACAG
 1651 GCACGGTCGG ACAGACTTGA CAGATCAGGC ATCAGGCCCT CTGCCTGGT
 1701 CCCGGCTCT TTAAGCAGGA ACGTGAATGG CCTCAAGATG TCTCACATGG
 1751 TCCCACACTG CTCACCCACC TGCTGCTGCC ATGAGGCACC TTGGGGCCTT
 1801 CTGCCCTTCC TTCCCTCTGTT CTTGGCCTT ATGTTCCCCG CCACCCACAGG
 1851 CCTTCCCCCG CCCCCCCCCCT CTGCAGACTT AGCCGTGCAT TGCAAGCATG
 1901 GAGGATTAAT CAGTGACAGG AAGCTGCGTC TCTCGGAGCG GTGACCAAGCT
 1951 GTGGTCAGGA GAGGCCAGC AGGGCCAGCC CCAGGAGTCT TTCCCAGATTC
 2001 TTGCTCACTG CTCACCCACC TGCTGCTGCC ATGAGGCACC TTGGGGCCTT
 2051 CCTCTTCCCTT CTGGGGGTCC TGGGGGCCCT CACTGAGATG TGTGGTGAGT
 2101 AACTGCCCTC TATCCTGTGC CTCTTCCCTC CTGGGTCTT AGTGGGGTGG
 2151 CTAGGGCTTCA GGATGAGGG ACTTACCTGC CTTCTAAGC TCCCACATGCA
 2201 GTTTGGCTT AGCTGGACCT CAGCATTAA CACATCCTAT TGTGATTGAT
 2251 TATATGTTTG ACTCCTCACC AGACAAGATC TCCGTTAATT CAGTCATTG
 2301 TTCACACATT CATTCAAGCCTC ATACTGAGCC TTTTCTGTGT CAGGCCAGT
 2351 GTTAGGCTTT GGGGAACGTG CAAAGCATGA GACAAGTCTA ATCCCTGCCA
 2401 TCCTAGAGCT TATGTTCTAG GGAAGGGGG CAGACAAAAG AAATGGTTAG
 2451 GTGCTCCAC CTGAAATCTC AGCATTTGG AAGGCTGAGG CGGGAGGGGA
 2501 GGATCGCTTG AGCTCAACAG TTCAAGGTCA GCCTGGGCAA CATAGGGAGA
 2551 CCCCCATCTC AAAAAAAATA AAAAAAAATT AAAAATAGCT GGGCATGGGG
 2601 AAGACTTTCT GAAAGACCAAG AGGACACATG GGAGCTGAA CTCGAAGGAA
 2651 GAAAAGGAGC TGGCAGGAAA GGAGTGGGG ACACACATTC TAGGCAGCAG
 2701 GAAGTGAGCC TTGGAGGTC CTGCCTGCTC CAGCTCTGTG CCCCAGGG
 2751 TCTCTGGAG CACAGTCTCC TGGGACCTGT CTATGAGTCT GAGCTTAGAG
 2801 GCTCAGGGCT GCTCCCTCAG ACAGGGAGGCA GAAGGAGAC TTTGGAACT
 2851 TTGGGGCCGC CACGGCCCTT TTCTCCTCTC CTGCACCTAG GATTACCTTG
 2901 AGCAATACAC TTTCACCCCCC ATGGTCTCTT GAGACCCCTGG GGAAACCCCTG
 2951 AGAGGTGGGT GCAGTCATGT CCAGGTGTCA AGTGAAGAAG TCGAGGGTTG
 3001 GAGGGGCTGA GTGACCCACT CAGGGTGCTC CACCTTTCC AGAGCTTGC
 3051 TGAACCTAGT TTTAGAACT TGAAGCCTCG TTTGTTTGTG TTTTGTGTTT
 3101 TGTTGAGAGA GGTCTCCCT CTGTTGCCA GGCTGGAGTG CAGTGGCAGC
 3151 ATCTTGGCTC ACTGCAGCCT CTGCCTGTG GGTCAAGTG ATTCCCCCAC
 3201 CTCAGCCTCC CAAGTAGCTG GAGACTGCAT GTGCATACTA CCATGCTTGG
 3251 CTAATTTTG TATTTTTTG TAGAGACAGG GTTTCGCCAT GTTGCAGG
 3301 CTGGTCTCGA ACTCCTGGGC TCAAGTGAAC CTCTTGCTC GGCCTCCCAA
 3351 ATTGCTGAGA TTACAGGCAGT GAGCCACCGT GCGCGGCCAG AACTCCAAGC
 3401 CTCTCATCTG TGTTCCATAA ATGCAATCAG ACACCTCAGG TCTGGGCCA

FIGURE 3A

3451 GGAACCCAG CTCTGGTC ATGTCCGGAC AGTCCCCAGG GGAGTTCTGG
 3501 GTTCAACCAG CAAGAGCTCT TCCTCCTGGC TGATCTGGTC CTCAGCCCTTG
 3551 GACAGTTAGT CCATTAACCT GACCCCACAG GAGCCCCAAT CCCTGGGGT
 3601 CTGGGAATC TTGAACCTGGG GTTTGGGGTG CAAATATCTG CACTGAGTC
 3651 CTTAATTGCA CCCAGCCTCA TTCTTTATC TGTAAGTGG GCTAAGAATG
 3701 CTCCCCGCGC TTCTCCTCG GTGTAGTAGC AGGAAGGATC CCATGACACC
 3751 TGCTCTCCA GTTAAAGCT CTATATGTAT GTTGTGAAT TGACAGGGAT
 3801 CGCTGCACAA ACGCTAATGC AAAGTGGGCT CCTGTGCTTC CTTTCTCTT
 3851 TCTTCTCTT TTTTTTTT TTAATTCT TCTAGAGATG AGGTCTCACT
 3901 ATATTGCCA GGGTTGGTT CAAACTCCTA GGGTCAAGCG ATCCCTCCAC
 3951 CTTGGCCTCC CAAACTGCTG GTATTACAGG CGTGAGCCAC TCTGCTGCG
 4001 TCCTATGCTT GTGAATGTCA ACAGCAATCA GCCCTTAGCT GGCAGGGCTG
 4051 GGTTGGTAGG GCGAGAGCTC ACCCAAGGCT GCTTTTATTA CCCTGCGTGA
 4101 ATCTGCTGG CCCCTTCTT CTAAGGAGGT TGCTCTGTG TTGTCAAGTCT
 4151 CTCCCCTTAC AGCTGGATAC TGATCTTCA GTTCTAACCT CTGTGCTGAC
 4201 TCATCGCTG GGAAGTGGAGA GCGGGGGTGC AGGTCAAGGGA ACTCCCTTGC
 4251 GCGTTCAAG AAAAGGGAAA AGGAAAGAGA GGTGAGGAGG GGGGCAGATG
 4301 ACCAGAGAGA CACAGGCTGA GAGAGACTGA GACAGACCA GAGAGCCTCA
 4351 CACATTGAGT GACAGAGACG GAGAAATGGA GATAGGCACC AAAAAATGGT
 4401 TCTCAGTGCAG AGAAAGGGAA AAAAGCAACC CCCCAGTCTC TCTTAACATC
 4451 TGGTAGAGAAA CCAGCCATGT GCTTTGGTCT GGGCCACAC AGCAAAGGAT
 4501 TATGTAGGT TTCACTGGG TGGATGGTC CTTTATAGCA ACAGGTATCT
 4551 GGGGCTGTCG GGAAAACAGA CACGAGGTTG TGGGACCCAG ACCCACAGAG
 4601 ATGGAGCTGT TCTAGGAGCT CTGGTCTCTT TTCTGGTCCC CTGGGATATG
 4651 GCACAGTGAAGG GGCACCATC AGGCAGCTGG AGCCCAGCAG CAACTGGGAG
 4701 GCAGTAAACA GGGACCGAAA GTGCAAGGTT ACCTCCGAGG CAAACTACTC
 4751 TAAGCTACCC TGTGCTGAGC TCAAGTCCCT TGGAAACTATC CCTAAGGCTT
 4801 CCGCTTCCAG AGTGTGTTGAG TATTTTCTT GCACAGCTTC GAATAAATCC
 4851 CACAGCAACA GTAAACCGC TGCAAGCTGT GACTGTTTC TAAGAGCTCA
 4901 TCTCACAAATC TCAGGTCTCTC TTCAATTAAA CAGAGATGGC AGGAAAGGCG
 4951 TTATTTGAG ATCTGCATGG AGGAAGTTC CACAGGAGCC TCAATTCA
 5001 AGCTGGAAGT TTGCGTTGTT TGGAAATTG ATGTGTAACA CGTTCTGCAT
 5051 GTGGGCTGAT GTTTTGAAACCGGTTAGCA CACACATTC GAAGGGCACC
 5101 AAAGAGCGGG GGCTTTGCAG TTAGGTCCAT CTTGGCTCT GCAGCCTTGT
 5151 GTAAGACATG ACACGACTTT GAACTTCTGT TTCTCTCTCT GTGCAAAGCA
 5201 ATGATGACAG TATCTACATC ACAGGACTGG CATGAGGACC AAGTGAAGATT
 5251 GGGCAAGGTG CCCCCGACA CCAGTCTCAC TGTCACTGCT GATGGGCAGA
 5301 GTGGTTGCTT GGCAGTAGCA TTCTCTATCT TCAGCCACCC ACCTCTCTT
 5351 CTGGCTCACT CCAACTGCTC TTAGAGATA CACGCTTCCC CTCTTTCTC
 5401 CTCCCACGTG CTTCACTGAT GGCTGCATTT CCCCCCTGAA GTGGTGTGT
 5451 GCTGGGTGGA GGTGGGGGTG AGGACATGTA TTCTCTGGAG AAGGCCCTGG
 5501 TAACGTCAAA GCACTTCTTT GCTGGTGGCC TGGCCCTGTG ACCTCATTTG
 5551 TACCATTTCTC TTCTCTAAGA AATACCAAGAG ATGGACAGCC ATCTGGTAGA
 5601 GAAGTGGGC CACCCCTCT TACCTTGAT GGACGGGCTT TCCCTGGAGC
 5651 ACTTGAACCC CAGCATCTAT GTGGGCCTAC GCCTCTCCAG TCTGCAGGCT
 5701 GGGACCAAGG AAGACCTCTA CCTGCACAGC CTCAAGCTTG GTTACCAAGCA
 5751 GTGCCCTCTA GGGTATTGCC ACACTCTCTT TTCCATGTC TTGCTCCACA
 5801 TACTAAGAGA TGGGAAACTT GGGTACTAGT TTGGGCCTGT CACCACTTG
 5851 TGGGAGACCC TTAGGCAAAT TTCTCCATC TATAGAATGG AGGACCTTTG
 5901 TCCATCTATA GAATGAAGGG GTGGGTTGGA TTAGATCAGA GATGCTAATG
 5951 CAAGGCTCTT TTTGCTACTA CTGTCCTATCA TGTCCTGAG GCAGACATAA
 6001 CTAATCCGTG ACTATACTCT TTGATGATGA GCCCCAGGAGC AGCATCTGAC
 6051 TCTATGCTCC CTTAGTGTGC CTGAGGCAGA TATCACTAAT CGATGACTGC
 6101 AGTCTCTAC ATTGAGCTTA GAAGCAGCAT CTGACTCTGT ATGCTCCCT
 6151 CCCATGCATG AGGCAGACAT CAGTAATCCA TGACCGCATT CTTTCATACT
 6201 GAGGCCAGAA GCAGCATCTT TTCTTTCTT TCCTCTCACT CTGTTGCCCA
 6251 GGCTAGAGTG CAGTGGCACA ATCTTGGCTT GCCCCAACCT CCAATTCCCG
 6301 GGTTCAAGTG ATTCTCGTGC CTCAGCCACC TGAATAGCTG GGATTACAGG
 6351 CGTGTGCCAC CATGCCACAG TGATTTTTGT ATTTTGTTA GAGATAGGGT
 6401 TTCACCATGT TGGCCAGGCT GGTCTTGAAC TCCTGACCTC AGGTGATCCG
 6451 CCTGCTTGG CTCCCCAAAG TGTTGGGATT ATAGGCATGA GCCACTGCAC
 6501 CAATCCAAA GCAGCATCTT TGTCCTCCCT TTTCAAGAGG CATCACAGAG
 6551 AGGCCCTGTT TGGGGTTGAGAAGGAGCC AAGAATCAGC CATGGAGTGC
 6601 CTCTTCTCA GACTCCCTCT TGAGAAGTGG GTGCAGGGGT GGAGAGAAAA
 6651 GAAGACTAGG CATAGTGGCT CATAACCTGTA ATCCCAACAT TTTGGGAGGC
 6701 TGAGGCAGGA AGATTGCTTG AGCTCAGGAG TTTGAGACCA GCCTAGGCAA
 6751 CATAGTGAGA CCACATCTCT TAAAAAAAG AAAAGAAAA AAAATGAGCC
 6801 AGGTGTAGTG ACTCATGCCT GTGGTCCCCA CTTCTCCGGA GGCAAAGGTG
 6851 GGAGGATCTT TTGAGGCTGA GAAATCGAGG CTACAGTGTGAG CCATGGTGGC

FIGURE 3B

6901 ACCACTGCAC TCCAGCCTGG GAGACAGAGA GACCCTATCT CAGTAAAAAA
 6951 AAAAATAAA AATATGGCTG GGTGTGGTGG CTCACGCCG TAATCCCAGC
 7001 ACTTTGGGAG GCCAAGGTAG GTAGATCACA TGAGGTTAGG AGTCGAAAC
 7051 CAGTCTGGCC AACATAGTGA AACCCCTGTCT CTACTGAAAA TACAAAAAAT
 7101 TAGCCAAGGG TGGTGGTGGG CAACTGTAAT CCCAGCTACT TGGGAGGCCG
 7151 AGGCAGAAGA ATCGCTTGAA CTCGGGAGGC GGAGGTTGCA GTGAGCTGAG
 7201 AACATGCCAC TGCACTCCAG CCTGGGAAAC AAGAGCGAAA CTCTGTCTCA
 7251 AAGAAAATAA ATAATAAAA TAAAAAAA AAAAAGGAGG GGGCATATGG
 7301 GTGAAGTATG GACAAGTAG TGGGGCAGGC ACAGATGATC TGGACACAGG
 7351 AGCCCTTGGA GTTATTCTT GAATCTAATC GTTCATCTT ATTAAATATT
 7401 TGTGGCATAC ACCTCACAAAC AACATAGCCA ACACACCTCC TTTTGGAGCT
 7451 TTTATCGAAG TTTCCACTG TTAAGATTT TTCCCGCTT GTGATGCGGG
 7501 TGGGGTGGGT GCTGTAAGCA GCCTTACGGG GTGGCAGTTT CTCACAAAGG
 7551 CATTAACTGG CCTTGTCTA GGTCCTGCCTT CAGCGAGGAT GACGGTGA
 7601 GCCAGGCCA GCCTCCATG GGCCAGCTGG CCCTCTACCT GCTCGCTCTC
 7651 AGAGCCAACT GTGAGTTGT CAGGGGCCAAC AAGGGGACA GGCTGGCTC
 7701 ACAGCTAAA TGGTCTCTGG AGGATGAGAA GAGAGCCATT GGTGAGCAGA
 7751 CACCATCCGC TGGGGTGGG GAGCAGCTGG GAGGGCTCAT CAGATGATAT
 7801 TCTCCAATGA GAATCAGAAC TTTGGGTTT CTCCCCAGGC GTCTTTCCC
 7851 CCATCCATTC TGCCCATCTC ACTGCCTACG TAGAGGCTCG AACCTGTCCC
 7901 CATAGCCATC CTTGACCCAG CTTTCCCGC GCTGCACACA TACTATTGAC
 7951 AGGTGTGTT CGTGGTTTTT TGTTTTTTGT TTGTTGTTT GTTTGAGTT
 8001 GGAGGTTGC TCTTGTGCC CAGGCTGGAG TACAATGGCG CAATCTCAGC
 8051 TCACCCCAAT CTCCTGCCTC TGGGTTCAAG CAATTCTCTT GCCTCAGCCT
 8101 CCTGAGTAGC TGGGATTACA GGCATGCGCC ACCACACCA GCTAATTGG
 8151 TATTTTGTAGT AGACGTGGGG TTTCTCCATG TTGGTCAGGC TGGTCTCGAA
 8201 CTCCGTACCT CAGGTGATCC GCTTGCCTTA GCCTCCGAAA GTGCTGGGAT
 8251 TACAGCCATG AGCCACTGCG TTAGGCCAAC TGACAAGCCT TGTATTGGCT
 8301 AGCCACCAAG ATTGACTTGA TTATCCACCT TCGGGACAAC TGGACAGCCT
 8351 GCTTATGACT TACCGCATAG TCTGTCTCTA CTAGCTCTC TGCCCTGACT
 8401 TGACCCAGCA TACAACAGCC AGAGCCAGGC TTTTCAATAT AAACCTGATC
 8451 TTGCTGGCAC TGCTTAAACC CTGCAGGGGC CTGCACACTGC TCCATGGCCC
 8501 AGCCTGTCTA CCCTTACCTT CTGCCCAGGC TCTGCTCATC CATTCTCTGC
 8551 CTCCCACACA CCTGCCCTCT GTGGGCTCCA GCCATACCAT CTCTCAACTC
 8601 ATAAGCCAGT TTTTTCATAC AGGCTCCCTC CATCTGGACT GGCTTCCCTG
 8651 CGTGCAGTTC ACTCTGTCTC TACCTTTGGC TCTGCCTCCA CCCATCCCTCA
 8701 GCCGCTCTCA GCATTACCTC CTTGGAGAAT CCTGCCTTGA CTTCCCAGCC
 8751 ACCCAAATAT CACTATTGG TCTGCAATTG CGTTGCAATT GCAGTCGCAT
 8801 GAGCAATTGC TGTGGTTGAG GCGCGAACTG CGCAAGTGCC TGTCTGCAT
 8851 GGGTCTCTG CTTCTCTAA GCACAGTGCC TGACACACAG TGAGACCTCA
 8901 GCACGTATGG GCTGAGGAA TGAAGGAATG AAGGATCCCA TGACCCAAA
 8951 GAGCCCTTTG GAAAGTCAG GCCAGGGTCC CAGGTGCTGG CGGGGCTGGC
 9001 TGCTGGTGTG GGGCAGAG GCAACCCCTC TGTTTTTTTC CCTCTCAGGG
 9051 CATGATCACA AGGGCCACCC CCACACTAGC TACTACCAT ATGGCCTGGG
 9101 CATTCTGCC CTGTGTCTCC ACCAGAAAGCG GGTCCATGAC AGCGTGGTGG
 9151 ACAAACTTCT GTATGCTGTG GAACCTTTCC ACCAGGGCCA CCATTCTGTG
 9201 GGTGAGTAGG TCAGACCGTG CCAAGGCCAG GCTGGCACTC CCTCAGTCCC
 9251 CAGGTCTGCA CTGATGACGT CCATACCCCTG GCCCCCACAC TCACCTTTCC
 9301 TTGGGCTCC TCCGAATCAA GTCCCTTAAAGG GACGAATTGG CGAGGGCTCA
 9351 TGGGTGATGC TCCAGCTGTG AGCCAGCTTT GGAGCTGTTA GGTGGATCTC
 9401 TTGAGGCCAG GAGTCAAGA CAACGTGGTG AAACCCCATC TCTACTAAA
 9451 ATAAAAAAAGT TAGCCGGCA TGGTGGCACA TGCCCTGTAGT CCCAGCTACT
 9501 CGGGAGGCTG AGGCAGGAGA ATCACTTGAA CCTGGGAGGC GGAGGGCTGCA
 9551 GTGAGTGGAG ATCCACCAAC TGCCCTCCAG CCTGGGCAAC AGAGTGA
 9601 AGACTCTGTC TCAAAAAATA AAAAATAAAAA TAAAACCTCC CTAGTGATTC
 9651 CAATGTCCAG CTAAGTTGG AAATAGGTGG TATGGGGTCA AGTCCCTTGT
 9701 GGCCCTCTC CTCCAGCTCT TCTCCCTAAC CTCTAGCCCT CAAGTTGCAG
 9751 AGTGTACAGC CAAACCAAGTT TGCCCTAAC TGAGCAGTT CCTGGGACAC
 9801 AGGATTTCAGA GAGTCCAGAC AAGGAAAGTC TTGGGCAAGAC CAGGTTGAGT
 9851 TGGTGCCTCTT AGCTGATCTG ACCATGTTGC CCTTCTCTC CAAGCCCTCC
 9901 TGTGGTTGTC CATAGCTACA AGGGCCTGAC CCTCAAGCCC CTGCCTGTCC
 9951 TGGCCCTTT GGCTCTCCAG CTCATTGCA GTTCTGTCCC CCACTTCAG
 10001 ACACACGAGC CATGGCAGGC TTGGCATTCA CCTGTCTGAA GCGCTCAAAC
 10051 TTCAACCCCTG GTCCGGAGACA ACGGATCACC ATGGCCATCA GAACAGTGCG
 10101 AGAGGAGATC TTGAAGGGCCC AGACCCCCCGA GGGCCACTTT GGGAAATGTCT
 10151 ACAGCACCCC ATTGGCATTA CAGGTGGGAA AGAGACCCCTG GAGCCATGGC
 10201 CACCCCTGGGG AACAGTCGGG TGGAGTGGTC AGGTGCTGGA ACACCTAGCC
 10251 CCTCCCTGCC GGCTGACCTC CTCTCTCTCT TCCTCACTCT ATCACCAGTT
 10301 CCTCATGACT TCCCCCATGC CTGGGGCAGA ACTGGGAACA GCATGTCTCA

FIGURE 3C

10351 AGGCGAGGGT TGCTTGCTG GCCAGTCTGC AGGATGGAGC CTTCCAGAAT
 10401 GCTCTCATGA TTTCCCAGCT GCTGCCCGTT CTGAACCAACA AGACCTACAT
 10451 TGATCTGATC TTCCCAGACT GTCTGGCACCC ACGAGGTAGC CCAACTTTTT
 10501 GTGGAAGCAC AGCCCTTAC AATCTGCTGC GCACCCATTG ACGTCCCAGT
 10551 GAGGGGAGGT TGCTTCATCC TGATTTGCTG AGTCAGCACA AGTTTGTGGG
 10601 TGTGCATGGG ACACAGTAGC CAAAATGTGG TCATAGCTTC TAGAAGCTCA
 10651 CAGTGTGGG AGGAAGACAG TAAATGGAGA TCCCTGGGCA TATCGCTTGT
 10701 GTGATACCCA GTACAGAAAT GTTTGGATGG ATGGATGGAT GGATGGATGG
 10751 ATGGATGGAT GGATGGATGG ATGAGGAGAG ACACATTG GTTAACCTCA
 10801 ATACACATG ATAAGCCCCA GTAGCAGCAT GATCCAGGCT TTCTCTGAGA
 10851 GAGGGTCTGA GGACGTGACT GGGATTGCCC AATTAAGAACAG
 10901 GCCAGGTGCA GTGACTCATG CCTGTAATCC CAACACTTTG GGAGGCCAG
 10951 GCGGGTGGCT CACCTGAGGT CAGGAGTTCG AGACCAGCCT GGCTAACATG
 11001 GCGAACCTCC ATCTATTAAA AATACAAAAA AGTAGCTGGG TGTGGTGGCG
 11051 AGTGCCTGTA ACCCCAGCTA AGCTACTCAG GAGGCTGAGG CAAGAGAAC
 11101 ACTGAACTC CAGAGTGGG CGTTGCAGTC AGCCAAGATC ATGCCACTGC
 11151 ACTCCAGTCT GGGTACAGA GTAAGACTAT GTCTCAAAAAA AAAAAAA
 11201 AAATGGAGAA GAAGGAAGCT GGACATGGTG GCTCGTGCTT ATAATCCTAG
 11251 CACTCTGGGA AGCTGAGGCA GATGGATTGC CTGAGCCCCAG GAGTTTGAGA
 11301 CCAGCCTGGG CAACATGGTG AAACCTGTC TTTACTAAAAA TACGAAAGAT
 11351 TAGGCCAGCA TGGTGGTAGA CACCTATAAT CCCAGCTACT AGGGAGGCTG
 11401 AGCCACAAGA ATCACTTGAA CCTGGGAGAC AGAGGTTGCA GTGAGGCCAG
 11451 ATCGCCCAT TGCACTCCAG CTTGGGCAGC AGTGTGAGAC TCTGTCCTCA
 11501 GAAAAAAACAA GAATGGATAG AGTGGAGGCA AGAAGAGGCA GGAAGAACAA
 11551 AGACACAGAG GTGCACAGAG TTTGGGGAA TTTTGAGGAA TGGTCTTGC
 11601 AAAGAGTGGG ATCTGGGAGA ATGAGTGGGA GTGAAAGCA GATGAATGAA
 11651 GAGAAGGTGA GCGCATCAGG GTAACAGAGA TGCCTTGTA ACAAAATGCAT
 11701 GTTCTAGGAA GAGCCCTCTG GAGTGTCTAGG TGCCAGAGAG GTGGGAGGAA
 11751 GGATACTGGG AGCAGAGAAA CCAGTGAGGG GCCTGATCTT GGGTGGTGGG
 11801 GAATGGAGGA CAGGGAGGCG CGGGATGGAA CCCAGGTGGT GGGGAATGAG
 11851 GGACAGGGGA GGCGGGATG GAAGCCAGGT TTCAGCTGAG CAGGTGGCGG
 11901 TGGCATTGAT GGAGATGAGG ACATGGGGAA GGACAAAGTC CAGGTGTCT
 11951 TGAGGGAAAGA CAAGAAGACA AATAATCCAG GCTCTCTGTC CTCACACCAG
 12001 CTGCCGCC CTTTCTTCCT GGCACAGTC TGTTGGAACC AGCTGCTGAG
 12051 ACCATTCTC AGACCAAGA GATCATCAGT GTCACGCTGC AGGTGCTTAG
 12101 TCTCTGCGG CCGTACAGAC AGTCCATCTC TGTTCTGCC GGGTCACCG
 12151 TGGAAGATG CCGTAAAGAAG GCCCCATGAGT TAGGAGGATT CACGTGAGAC
 12201 TCCCCACCTC CAGTCCTCAC CCCACCCAAAC CTCACATGCC TGATAACAGG
 12251 GTCACAGAAA AGACGGGGAA CAGAGGAGAG GTTCCCTCG GGAGAGACAC
 12301 TGGCCCTGCT TCTGCTTCTA CCTGCTCAGC TCCTTCTTG CCCACGGTGT
 12351 TATGGAAACA GGGAGCCATA GGCCAGCATT GTCACTGAGA GAGCAGGCTT
 12401 TGGAGGCAGA GCCCCCAAGT TGGAAATCCCA ACTCTAACCA GCTAGGTTCC
 12451 AGGTAGGCAC CCACAATTCA CCGAGGAGAA CAGTTGTGCC CTTCCCTGC
 12501 AGGGCCACTG TGAAGAGTCAAG AGGAGTTAGT ACACATAGAG ATAGTGGCAT
 12551 GTGCTTTTA TATGTGCAAG GTCCAGCACA TAGCAAGCGC TCAACACAGC
 12601 GTTGCTTTCA TCAGAGTAAG AACTGTTTT TGTTTGTGTT TTTGTTGTT
 12651 TTTAAGAGAC AGGGTCTCAA TCTTATCACC CAGGCTGGAG TGTAATTGTG
 12701 CAATCACGTC TCACTGCAGT CTCGAACCTCT GGGGATGAAG CAACCCCTACT
 12751 GTCTGCTCTC AGCCTCCAA ATAGCTGAGA CTATAGGCAC GTGCCACACAA
 12801 ACCCTGGTA ATTTTTTTTT TTTTTTTTTT GAGATAGGGT CTCTGCTGT
 12851 TGCCCAGGT GGTCTCAAAT TCCCTGGCCTC AAACCATCCT CACACCTGAG
 12901 GCGCTCAAAA TATTGGGATT ATAGGTGCGA GCCATCATGC TCAGCCAGAA
 12951 TAATAACTGG TTTTTTTTGT TTTTTTTTGT AGACAGAGTC TCACTCTATT
 13001 ACCCAGGCTC TGGAGGCCA ACTCGTGTGTT GTGTATTGTT TTATTGTTAT
 13051 TTATTTATTT ATTTCGAGAC AGAGCCTCTC TCTTTCACCT AGGCTGGAGT
 13101 GCAGTGGCCG AACATCTGCCT CACTGCAACC TCCGTCCTCC GGGTCAAGT
 13151 GATTGTCTG CCTCAGGCTC CTGAGTAGCT GGTGCTACAG GCGCGTGCCA
 13201 CCATGCCAGC CTAATTGTTG TATTGTTAGT AGAGACAGGG TTTTACTATG
 13251 TTGGCCAGCT GTTCTAAAC TCCCTGAACCTC GGGTGATCTG CCTGCCCTCG
 13301 CCTCCCAAAG TGCTGGGATT ACAGGCATGG GCCTCCGTGC CGGGCCATGT
 13351 ATTTATTAG GCAAGGTCTC TCTCTGTTAT CCAGGCTGAA GTGCAGTGGC
 13401 ACATTCTAG CTCACTGCAG CCTCAAATTAA TCCAAGTAAC AGGGACTACA
 13451 GGCATGCAACC ACCACACCCA TCTACTTTTT TTTGAGATGG AGTCTCCCTC
 13501 TGTCGGCCAG ACTGGGTTGC AGTGGCACAAT TTTCAGCTCA TGGCAGCATE
 13551 TACCTCCAG GTTCAAGCGA TTCTCCCTCC TCAGTCTCCC GAGTAGCTGG
 13601 GACTATGGGC ATGCACCAACC ATACCTGGCT AATGTTTATA TTTTGAGTAG
 13651 AGATGGAATT TTGCCATTGGCAGGCTG GTCTTGAGCT CTTGACCTCA
 13701 AGTGATATGT CTGCCCTCAGN NNNNNNNNNNNNNNNNNNNNNNNNNNNNN
 13751 NNN

FIGURE 3D

FIGURE 3E

17251 NNNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 17301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 17351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 17401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 17451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 17501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 17551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 17601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 17651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 17701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 17751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 17801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 17851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 17901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 17951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 18001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 18051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 18101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 18151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 18201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 18251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 18301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 18351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 18401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 18451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 18501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 18551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 18601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 18651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 18701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 18751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 18801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 18851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 18901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 18951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 19001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 19051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 19101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 19151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 19201 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 19251 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 19301 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 19351 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 19401 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 19451 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 19501 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 19551 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 19601 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 19651 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 19701 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 19751 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 19801 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 19851 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 19901 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 19951 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 20001 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 20051 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 20101 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN
 20151 NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNNNNNNNN NNNCAAATC
 20201 AACAGTTGC ATAAATCACT CCTCTATCTT CCTTGGGTG GAAAGTGGAT
 20251 GGGAGTTATA ATTGAGTTC TCTTTGTCT TAGTCCATTG AAGCTGCTAT
 20301 TACAAAATAC CATAAACTGG GTGGCTTATA AACAGCAGAA ATGAGGCCGG
 20351 GTGCGGTGCG TCATGCCAT AATTCCACCA CTTTGGGAGG CCAAGGCAGG
 20401 TGGATCACCT GAGATCAGTA GTTCAAGACT AGCCTGACCA ACATGGTGAA
 20451 ACCCTGTCTC TACTAAAAAT ACAAAAAAATT AGCTGGGGGT GGTGGCGGGC
 20501 ACCTGTAATC CCAGCTACTC AGGAGGCTGA GGCAGGAGAA TCGCTTGAAC
 20551 CCAGGAGGCG GAGGTGCGG TGAGCTGAGA TCACGCCATT GCATTTCAGC
 20601 CTGGGCACAA AGAGTGAAAC TCCATCTAA AATGAAATAA AATAACAGAA
 20651 ATGTATTTCT TAACAGTTCT GGAGGTGGG TGGGCAGTCC CAGATCAGGA

FIGURE 3F

20701 CACTGACAGA TTCAGTGTCT GATGGGGGCC CACTTCTGG TGTTACCTGC
 20751 TGGCTGTGTT CTCACATGGT GGAAGGAACA TGGCAACTTT CTGGGGCCTT
 20801 GTTTTTAAAT TTAAAAAAA AAAATATTT CCTGGCCCTT GCCTGCTGAA
 20851 GGAACCTCTT TTATAATGGT ACTTAAAAAT TTTTTTTTT GAGATGGGG
 20901 TCTCACTCTG TCACCCACGC TGAGTGCAGT ATCACAATCT CAGCTCACTG
 20951 CAACCTCTGC CTCCCTGGCT TAAGCGATCC TCCCACCTCA GCCTCCTGAG
 21001 TACGTGTGAC CATAGGCCA TGGCACAAAG CCCAGCTAAT TTTTGTTATT
 21051 TTTAGTAGAA ATGTGGTTT ACCATGTTGC ATAGGCTGGT CTCGAACCTC
 21101 TGAACCTAACAG TGATCTGCCT GCCTTGGCCT CCCAAAGTGC TGGGATTCTA
 21151 GGTATGAGCC ACCCTGCTCG GCCTATAATG GCACCTTCCT ATCCCATTGA
 21201 TGAGGCTCTA CTCTCATGAC CTAATCATCT CCCAAAGGCC CTAAGGCCCTC
 21251 CTGATACCAC CACCTTGGG GTTAGGTTTT AACATATACA TTTTGGGGG
 21301 ACACAGACAT TTTAGACCAT AGCACCTCCA TTGAAAGGAA ACATTCTGA
 21351 CACCTGGCTA TCTCAAAGGG CCCTTTCACT TCCCCCTGCAG GCTGCATTCC
 21401 CACATCACCA ACAAGAGCAG CGACACTCAG TCAGAGGTTA AATAACTTGT
 21451 CCAGAGTCAC AGCAGTAATG ATGACAGAG CTGGGGCTTG AATCCAGGGC
 21501 TCCTCCTAGA GCCTGGATTG TGTGTAGTG GTGAAAGCTG ACTCCTGGGA
 21551 GACTTCTCGG TGGTCCCTGGT TCTCTCTCCA GACTGCACTG CGCAAGTTTC
 21601 TCTTCCTGAT GGTCCCTAGG GTATTACAAA GACAGTGGCC CTGCTGTCA
 21651 GGTGTTTTTA TTACCAAGATG AGGTCAATGGC CTCAGGAACC CTGTAGGAAG
 21701 CTGAGTTCTAG AGTCTTTAGG CAGGCTTTAG GGAGGTTCCA GCTTCCCACC
 21751 ACCAACCCCCC AGGTGGATTG TTACAGACTC TAGCCTCAGG GTGGGGGGTC
 21801 TGGAAGATGA GGTGCGGGG TGCGATATTG TGCCCAATTG GCCCCCTCCTT
 21851 GCTCAATCTG TTTCTGCAAGG TATTGCTGAC TACAGACCA AGGATGGAGA
 21901 AACCAATTGAG CTGAGGCTGG TTAGCTGGTA GCCCCCTGAGC TCCCTCATCC
 21951 CAGCAGCCTC GCACACTCCC TAGGCTTCTA CCCTCCCTCC TGATGTCCT
 22001 GGAACAGGAA CTCCGCTGAC CCTGCTGCCA CCTCCTGTGC ACTTTGAGCA
 22051 ATGCCCTCTGG AGATCACCCCC AGCCACAAAGC CCTTCGAGGG CCCTATACCA
 22101 TGGCCCCACCT TGGAGCAGAG AGCCAAGCAT CTTCCCTGGG AAGTCCTTCT
 22151 GGCCCAAGTCTG GGGCAGGCTG GGGCTGAGG TCTCCCATGAG AGGCCACCCCC
 22201 ATGGTCTGAT GGGCATGAAG CATCTCAGAC TCCCTGGCAA AAAACGGAGT
 22251 CGCGAGGCCG CAGGTGTTGT GAAGACCACT CGTTCTGTGG TTGGGGGTCT
 22301 GCAAGAAGGC CTCCCTCAGCC CGGGGGCTAT GGGCCCTGACC CCAGCTCTCC
 22351 ACTCTGCTGT TAGAGTGGCA GCTCCGAGCT GGTTGTGGCA CAGTAGCTGG
 22401 GGAGACCTCA GCAGGGCTGC TCAGTGCCTG CCTCTGACAA ATTAAAGCA
 22451 TTGATGGCT GTGACCTGC TACAGTGGCC TGGTGCCTCA TACTCTCAG
 22501 GTGCAGGGGC AGGGACAAGA GAAGGGGGAA GTAACCCCCAT CAGGGAGGGAG
 22551 TGGAGGGTGC CTGAGGCCGC ATGTGGGCAT TGGGGGAGTG ATGGGAATGC
 22601 CAGCAGTGTAG CAGCTTGACT ACTGACTGAG CACCCACTAC TATGACTGAG
 22651 CACTCACTCG CTAGATACTA TCTTGAACTG CTCTGTGAGG TTGTTGATAT
 22701 TTTCATTTTT ATCTGTGTT TACAAATCG GAAACTGGGA GCCCAGGGCGT
 22751 GGTGGCTCAC GCCTGTAATC CCAGCACTTT AGGAGGCCA GGCAGGTGGA
 22801 TCACAAGTC AGGAGTTGA GATCAGCCTG GCCAACATGG TGAAACTCCA
 22851 TCTTCTAGAA AAATACAAA ATTAGCCAG GCATGGTGTG TTGTTGATAT
 22901 ATGCCTGTAA TCCCAAGTTAC TTGGGAAGCT GAGGCAGGGAG AATTGCTTGA
 22951 ACCCTGGAGG CGGAGGTGT AGTGGAGCCG GATCACGCCA TTGCACTCCA
 23001 GCTTGGCAA GAAGAGAAAC ACTCTCAAAA AAAAAAAAAA ATCAGGAAAC
 23051 TGGTGCTCAA AAAGGAAAG TGACTCACCA AGGTACACAGA CTAGGCAGTG
 23101 ATGCTGGGG AACCTGGCTC AGGGGACACA GACCTGGCCT GGGGCAGCCT
 23151 TGCAGCTCCT CCATAAAAT ACTGAAAATG AGGGGCTTCG ATGATGGTTA
 23201 TAATCGGTATG CGAGAGCCC AACTCAACTG GAGCCCTGGG ACCCAGAAGC
 23251 TAGGGTCTCA CTCCCTGCTT TTCCACAAGG CACCATTAAGG GCATCACCCCC
 23301 AGGCCTCGGC AGCCACGAGC CAGGGATCCT GCCTCTCATT GGTTGGGGC
 23351 TTAGGGCTC TGGGCTGCC TCTTGAAGAG GGGGTTCAAGC CCAGCGAGGC
 23401 ACCCCCTATG CTGCACCCCA CCAAGGTTAG GAAGAGGTCC TGTCCTCAGT
 23451 GGGGCCCCCT GATGAACAGC CCATCAGGTC TGCGTCCACA TGCCTTGGAA
 23501 GAGATGGTGA CATACTCAAA GTCCCTGAG CCGCATATTA AACCACTAG
 23551 AGCACCATCT TCAAAACATT AGGGTCTGAG AAGATAGGGG AAGTAAGCAA
 23601 TTTAAACAT TTCTTTATAT TGGGCCAGGT GCAATGGCTC ACGTCTGTAA
 23651 TCCCAGCGCT TTGGGAGGAC GAGGATCACC TGAGGTCAAGG AGTTCAAGAT
 23701 CAGCCTGGCC AACATGGAGA AACCCCATCT CTACTAAAA TACAAAATT
 23751 AGCTCAGGGG TGGTGATGTG CACCTGTAAT CCTAGCTATT CAGGAGGCTG
 23801 AGGCACAAAGA ATTGCTTGAG TCAATATTGC ACCACTGCACT TCCAGCCTGG
 23851 GCAACAGCGA GACTCTGTG TCAAAAAAAA AAAAAGATAT TTGCTGAAAA
 23901 GACCCAGCCT GCCAAACTCA GGGGCAGCCA AGGGAGGTAG TGAAATGGAA
 23951 GTTGGAGCTC AGCGCTCCCA CACCTCCACT GCCCTCAGGC CTTCTCTGCC
 24001 TCTTCTCCAT CAGTCAGCTG CTTCTGGCA TGGTCCCTGGC AGAGACTTGG
 24051 CCTCCTTCCA GTTCAAGCTC CCTCTTAGAT TGTGTCCAC GCCACTGAGT
 24101 CTTTGGGACA CTGGGTCAAGA TGTCTAGTCT GGCACAATTG GCAGGAATCC

FIGURE 3G

24151 CAAGAACAG TGTGAGTGA GGGACAGTCG TGTTGAGTGC CCTCCATCTG
 24201 GGACTGGAG GCAGGTCTAT GTCAGGCCCTG CATTAGATC TCTAATGGCT
 24251 CCAGACAAGC CCCCTCAGCT CACTAAGCCT GTTTCCTAAC ACAGCTGTGG
 24301 GATGGTGCTT TGGTTTACAT AGCACCGCAT ACCATCATAG ATCACATGGG
 24351 GAAACTGAGG CCCCCAGGAGT GATCTGCTGG CACATGCAGT GACAAGAGGA
 24401 GAGGCCATC TCAGCCTTGC AGCAAGGTTG CCAGAAATCG ATTCTCGCCC
 24451 CCATCCCGTA AAGATAGCTG GGATTACAGG TGTGCACCCAC CATGCCAGC
 24501 CTAATTCTTG TATTATTAGT AGAGATGGGG TTTCACCATG TTGTCAGGC
 24551 TGGTCATGAA CTCTGACCT CAAGTGATCC ACCCGCTTTG GCCTCCAAA
 24601 GTGCTGGAT TACAAGCATG AGCCACAGTG CCTGGCCTGA CCCTGCTCTT
 24651 TTGAAAGACC ATTCCCCAA ATTCTGTGCA CCTGTGTGCC TTTCTCTCT
 24701 CTGCCTCTC TCAGCTCTGC CCCGCTCTCC TCCCTCTCC TCTGGCAAAT
 24751 CCCACTCATC TCTTGAAGCC CTTCTTCCAG GGGAAAGCCCT GATCATGCTG
 24801 CTTTCTCTG TGGGAGGGAT GAAGGACGTG GCCCACGGAG TTTGTTTGT
 24851 TTGTTTGA GATGGAGTT TGCTCATGTT GCCCAGGCTG GGGTACAATG
 24901 GTACGACTC AGCTCACTGC AACCTCTACG TCCCCGGTTA AAGCGGTTCT
 24951 CCTGCCTTAG CCTCCCCAGT AGCTGGGATT ACTGGCATGA ACCACCCAC
 25001 CTGGCTAATT TTGTTTTT AGTAGAGATG GGGTTTCTTC ATGTTGGTCA
 25051 GGCTGGCTC GAACTCCAA CCTCAGGTGA TCTGCCTGCC TCGGCCTCCC
 25101 AAAGTACTGG GATTACAGGG TTGAGCCACT GTGCCTGCC CAGGCCACG
 25151 GAGTTTAAG AGGCTTCTG TGGCAGTGGC ATCCAGACGG AGTGCAGAAA
 25201 CTCAAAGTG AAGGCCAAG GCTCAGGGAA GGGGGAGTGT GAGTTGAGGA
 25251 GTCTCTTGGC TGCCAGGGCC AGAAACCGAA CTCCAAGCCT CTCCACAACA
 25301 GCGGGTGTAG AGCATGTAGA ATCAGAGAGG AGGCTGAGCC ATGCAGCCCC
 25351 GAGAAGAGGG GAATGCCACT GAGCCACAGA GACCCAGTGC CACTGCCAGG
 25401 TGTCTCTGCC TCCACTTCCC ATGACCCGGC CTGTCCTGT ATGCAGGCTT
 25451 CACCCCTCTCT CGTTGTACAT TGTACACATT CTAGGTGACA CCAGCAGCTT
 25501 CTGATTCTCA TCTCCCATAA CATCAGCCCC CCAGAGAGGG GACAACGTCT
 25551 GAGCTGATAA CATAATAGAT GCCCCCTTCC TGGAGGCCAT GGTCACTGGTC
 25601 AGCGTGGAGA GGATGAAGGC TGAGCAGGCC GGATCGGGGG TCTAGAGGGG
 25651 AAGGAGGTG AAGTTGAGAT CACAGACCTG TGGTCAGGTG GCCTGGGAAG
 25701 GGTTTGACGA GTGTCGGCCC AAAGAGCTTG GAAAGGGATTT TGCTGCTGTG
 25751 GGTGAGCACT GCCTCTCCCC TTAGGGACAA CAGCCACCTC TTCTCTCCCC
 25801 ATTTGCCTTT CCCTCTGTA GATATGAAAC ACAGGCCCTCC TTGTCAGGCC
 25851 CCTACTAAC CTCCTGATG GGGAAAGCGG CCGGAGAAAG GGAGTTCTGG
 25901 CAGCTCTCC GAGACCCAA CACCCCCACTG TTGCAAGGTC AGTCATGGCC
 25951 TGACACTCTG GATGTTGCTC TACCCCCAAG CTTACTCAGC CAAGAGGCTT
 26001 CATCAACTCA CCCCAGCTTT CCCTAGCACC CTCCCTGGCC ACACCTTCAC
 26051 AAAATCACTG ATGCTCAAG TTGGATATAA TATATTGAAAC TGAAGCCTTA
 26101 GCATTTTAT GCAAGTTACT GTGGAAATTG TAGGAAACCA GACAGATTAC
 26151 AAAAAAAA AAAAAGTAGA AGAAAATTAA CATCACCTAG GATATACTAC
 26201 CTAGGAAATA CGTCTTTAT TTTGAGATGG AGTTTCGCTC TTGTTGCCA
 26251 GGCTGGAGTG CAGCGGTG ATCTCGGCTC GCTGCAACCT CCGCCCTCTG
 26301 GGTTCATGTC ATTCTTCCAC CTCGGCCCTTC CTAGAGCCCA AGTGGCTG
 26351 CTGCCCTGCTC CTCCCAAAGT TCTGGGATTA CAGGCATGAG CCACCGCACC
 26401 CAGCCAAAT TACTTAACTT TTCTTCTAGA TACTTTTAA AAATATGGCA
 26451 GTAAGTTTT CATAAAAAAT GGAGCCATGC TATCCAGTGG AAATTAAATG
 26501 TTGCCACAT GTATAACTTA AAAATTTCAT ATATGTGTAT ACATATATAT
 26551 GAAATATATA TATACAGACA CACATATATA TGTATACATA TATATACACA
 26601 TATATATGTA TACATATATA CACACATATA TGTATACATA TATATACACA
 26651 CATATACACA TATATACACA CACATACATA TATACACACA CATATATACA
 26701 CACATATATA CACACATGCA CACATATATA TGTATACATA TATACACACA
 26751 TGTATACGTA TATATACACA CATATATACA CACATATATA TACACACATA
 26801 TACACACATA CACACACATA TATACACACA TATATACACA CATATATACA
 26851 CACATATATA TGTATACATA TATATACACA CATATATACA TACACACACA
 26901 TACATATATA CACATATACA CATATACACA CACATATACA CACATGTATA
 26951 CATATATATA CACACATGTA TACATATGTA TACACACACA TATATGTATA
 27001 CATATATACA CACATACATA TGTGTACATA TATACACACA TACATATGTA
 27051 TACATATATA CACACAT

(SEQ ID NO:5)

FIGURE 3H

Isoform 1:

FEATURES:

Exon: 2031-2094
 Intron: 2095-5569
 Exon: 5570-5762
 Intron: 5763-7571
 Exon: 7572-7741
 Intron: 7742-10000
 Exon: 10001-10173
 Intron: 10174-10298
 Exon: 10299-10485
 Intron: 10486-12027
 Exon: 12028-12193
 Intron: 12194-25821
 Exon: 25822-25939

Allelic Variants (SNPs):

DNA Position	Major	Minor	Domain	Protein Position	Major	Minor
921	C	T	Beyond ORF(5')			
1781	C	T	Beyond ORF(5')			
1850	G	A	Beyond ORF(5')			
2839	A	G	Intron			
3730	G	A	Intron			
6631	G	A	Intron			
6945	-	A	Intron			
6952	A	T	Intron			
7457	G	A	Intron			
7830	T	A	Intron			
8089	T	C	Intron			
8551	C	T	Intron			
9269	G	C	Intron			
9362	C	T	Intron			
9782	G	T	Intron			
11493	G	A T	Intron			
12260	A	G	Intron			
13086	T	C	Intron			
13183	T	C	Intron			
21240	C	G	Intron			
21695	A	G	Intron			
22058	C	T	Intron			
22233	C	A	Intron			
22245	C	-	Intron			
22375	C	T	Intron			
23042	A	- T	Intron			
23344	T	C	Intron			
23873	A	-	Intron			
24764	G	T	Intron			
24939	T	C	Intron			
24945	G	A	Intron			
25092	C	T	Intron			
25428	T	G	Intron			
25513	C	T	Intron			
25684	C	T	Intron			
26165	A	-	Beyond ORF(3')			

Context:

DNA

Position

921 TTGGAGATATTTAAGGTATAGTGTCTTCACAAATTGAGCTGAAAGGAACTGTTAGGA
 TGATCTTGCCTAACCCCTCATCTCACACAGGAAGAACTATTTAAACTCGAGAGGTTAA
 GTGACCTGGCCAAGTCACACAGCCACCACTAGTTAACTCGTATACATTGATTCTCCTGT
 GGGGCTGGGCAGATGAGGAATCTTGTCTTCCCTGTTGCAGAGATTTTTTGAG
 GTTACTTCCGAGTTCTGGCAAGTACCCCTGCTTCTGGTAGCTTGTCTCGATTCAAT
 [C,T]

FIGURE 3I

TCATTCTTTTATTTTATTTGAGACAGGGTCTCACTTGTCAACCAAGCTGGA
 GTCAGTGGTGTAACTTGCTCACTGTAGCCTCACCTTGGGTTCAAGCGATCCTCC
 TGCTCAGCCCCAAGTAGCTGGGATTACAGACGTCGCCACACGCGAGCTAATTAA
 TGGTTTTTGTATGTGTTTTGTGTTAGAGACAGTGTCCCCATGTTGCCAG
 GCTGGTCTCAACTCCTGAGCTCAAGTGTATCTGCCCTCAGCCTTCAAAGTGTAGG

 1781 ACAGCTGACTCCACAACTGCTGTCACGTGACCACTGCAGCTGCAGCTCCCCTTCCACTC
 CTTGCTCGGCTAGGGCACTACCAAGGGCTCCTTGGTAAGGAGTACCGGGTAGGC
 ACCCGGTCTGCCAACCACTGGAACAGCTGGGGGACAGCAGACAGGCACGGTCGG
 ACAGACTTGAAGATCAGGCATCAGGCCCTCTGCGCTGGTCCCAGGCTCTTAAGCAGGA
 ACGTGAATGGCTCAAGATGTCTCACATGGTCCCAGTGCCTCCCTTGTGTTCC
 [C, T]
 TACCTCCAGGAGGGCTGCTGCCCTCCTCTGTTCTTGGCCTATGTTCCCCGC
 CACCCACAGGCCCTCCCCCCCCACCCCTCTGCAGACTTAGCCGTGCATTGCAAGGCATGG
 AGGATAATCAGTGAAGGAAGCTGCTCTCGGAGCGGTGACCAGCTGTGGTCAGGAG
 AGCTCAGCAGGGCAGCCCCAGGAGCTTCCCAGTCTTGCTCACTGCTCACCCACCT
 GCTGCTGCCATGAGCACCTTGGGCCCTCTCCCTGTTCTGGCCTATGTTCCCCGCC

 1850 GCCTAGGTGGCACTACCAGGGGCTCTTGGTAAGGAGTACCGGGTAGGCACCCGGTCC
 TGCAATCCACCACTGGAACAGCTGGGGGACAGCAGACAGGCACGGTCGGACAGACTTG
 ACAGATCAGGCATCAGGCCCTCTGCCCTGGTCCGGCTCTTAAGCAGGAACGTGAATG
 GCCTCAAGATGTCTCACATGGTCCCAGTGCCTCCCTCCCTTGTGTTCCCTACCTCCA
 GGAGGGCTGCTCTGCCCTCCCTGTTCTGGCCTATGTTCCCCGCCACACAG
 [G, A]
 CCTTCCCCGCCACCCCTCTGCAGACTTAGCCGTGATTGCAAGGCATGGAGGATTAAT
 CAGTGAAGGAAGCTGCTCTCGGAGCGGTGACCAGCTGTGGTCAGGAGAGCCTCAGC
 AGGGCAGCCCAGGAGTCTTCCCAGTCTTGCTCACTGCTCACCCACCTGCTGCTGCC
 ATGAGGCACCTTGGGCCCTCCCTCTGGGGCTCTGGGGCCCTACTGAGATG
 TGTGGTGAAGTAACCGCCTATCCTGTGCCCTTCTGGTCCCTAGTGGGTGG

 2839 AACATAGGGAGACCCATCTCTACAAAAAATAAAAAAATAAAAAATAGCTGGCATGG
 GGAAGACTTTCTGAAGACCAAGAGGACACATGGGAGCTGAAACTCGAAGGAAGAAAAGGA
 GCTGGCAGGAAGAGTGGGGACACACATTCTAGGCAGCAGGAAGTGAAGCTGGAGG
 TCCCTGCCTGCTCCAGCTGTGCCAACGGGTCTCTGGAGCACAGTCTCTGGACCT
 GTCTATGAGTCTGAGCTAGAGGTCAGGGCTGCTCCCTAGACAGGGAGGCAGAGGCAG
 [A, G]
 CTTGGGAACCTTGGCCGCCACGCCTTCTCCTCTGCACCTAGGATTACGTT
 GAGCAATACACTTCAACCCCATGGTCTTGTAGACCCCTGGGAAACCCCTGAGAGGTGGG
 TGCAGTCATGTCAGGTGCAAGTGAAGAAGTGAAGGGTGGAGGGCTGAGTGAACCCAC
 TCAGGGTGCCTCACCTTCCAGAGCTTGCTGAACCTAGTTAGAGACTTGAAGCCTC
 GTTGTGTTCTGTTGTGTTGTGAGAGAGGTTCTCCCTGTTGCCAGGCTGGAGT

 3730 GACACCTCAGGTCTGGGCCAGGAACCCAGCTTGGTCATGTCGGACAGTCCCCAG
 GGGAGTCTGGGTTCAACCAAGCAAGACCTTCTCTGGCTGATCTGGCTCAGCCTT
 GGACAGTTAGTCATTAACTGACCCACAGGAGCCCCAACCTGGGCTGGGAAT
 CTTGAACGGGTTGGGTGCAAATATCTGCACTGAGTCAGTAAATTGACCCAGCCTC
 ATTCTTATCTGAAAGTGGCTAAGAATGTCCTGGCTCCTCTGGTGTAGTAC
 [G, A]
 AGGAAGGATCCCAGACACTGCTCTCCAGTTAAAGCTATATGTTGAAAT
 TGACAGGGATGCTGCACAAACGCTAATGCAAAGTGGGCTCTGTCCTTCTCTT
 TCTCTCTCTTTTTTTAATTCTCTAGAGATGAGGTCTCACTATATTGCCCA
 GGGTTGGTTCAAACTCTAGGGTCAAGCGATCCTCCCACCTTGGCTCCAAACTGCTG
 GTATTACAGGGTGTGCCACTCTGCTGGCTCCTGTTGTGAATGTCAACAGCAATCA

 6631 TGAATAGCTGGATTACAGGCCTGTCACCATGCCAGCTGATTGGTATTTGGTA
 GAGATAGGGTTTACCATGTTGGCCAGGCTGGTCTTGAACCTCTGACCTCAGGTGATCCG
 CCTGCTTGGCTTCCAAAGTGTGGGATTAGGCATGAGGCCACTGCACCAATCCAAAA
 GCACCATCTTGTGCTCCCTTCAAGAGGCATCACAGAGGCCCTGTTGGGTTGA
 ATGAGAGGCGAAGAATCAGGCATGGAGTGCCTCTTCAGACTCCCTTGTGAGAAGTGG
 [G, A]
 TGCAGGGGTGGAGAGAAAAGAAGACTAGGCATAGTGGCTCACCTGTAATCCAAACATT
 TTGGGAGGCTGAGGAGGAAGATTGCTTGTAGCTCAGGAGTTGAGACCAGCCTAGGCAAC
 ATAGTGAAGACACATCTTAAAGAAAAGAAAAGAAAAGAAAATGAGCCAGGTGAGTGA
 CTCATGCCGTGGTCCCCACTCTCCGGAGGCAAAGGTGGGAGGATCTTGTGAGGCTGAG
 AAATCGAGGCTACAGTGAGCCATGGTGCACCACTGCACTCCAGCCTGGGAGACAGAGAG

 6945 AGAAAAGAAGACTAGGCATAGTGGCTCACCTGTAATCCAAACATTGGGAGGCTGAG
 GCAGGAAGATTGCTTGTAGCTCAGGAGTTGAGACCAGCCTAGGCAACATAGTGAAGACCAC
 ATCTCTTAAAAAAAGAAAAGAAAAGAAAATGAGCCAGGTGAGTGAACATGCTG

FIGURE 3J

TCCCCACTCTCCGGAGGCAAAGGTGGGAGGATCTTGAGGCTGAGAAATCGAGGCTAC
 AGTGAGCCATGGTGGCACCACTGCACTCAGCCTGGAGACAGAGAGACCCATCTCAGT
 [-, A]
 AAAAAAAAAAATAAAATATGGCTGGGTGGCTCACGCCTGTAATCCCAGCACTTT
 GGGAGGCCAAGGTAGGTAGATCACATGAGGTTAGGAGTCGAACACCAGTCTGGCCAACAT
 ATGAAACCCCTGCTACTGAAATACAAAAAATTAGCCAAGGGTGGTGGCAACT
 GTATCCAGCTACTGGAGGCCAGGCAAGAAGATCGCTGAACTCGGGAGGCCAGG
 TTGCACTGAGCTGAGAACATGC

 6952 AAGACTAGGCATAGTGGCTCATACCTGTAATCCCAACATTGGGAGGCTGAGGCAGGAA
 GATTGCTTGAGCTCAGGAGTTGAGACCAGCCTAGGCAACATAGTGAGACCACATCTCTT
 AAAAAAAAGAAAAAGAAAAAAATGAGCCAGGTGAGTGACTCATGCCGTGGTCCCCAC
 TTCTCCGGAGGCCAAGGTGGAGGATCTTGAGGCTGAGAAATCGAGGCTACAGTGAGC
 CATGGTGGCACCACTGCACTCCAGCCTGGGAGACAGAGAGACCCATCTCAGTAAAAAA
 [A, T]
 AAAAAAAATAATGGCTGGGTGGCTCACGCCTGTAATCCCAGCACTTTGGGAGGC
 CAAGGTAGGTAGATCACATGAGGTTAGGAGTCGAACACCAGTCTGGCCAACATAGTGAAA
 CCCTGTCTACTGAAATACAAAAAATTAGCCAAGGGTGGTGGGGCAACTGTAATCC
 CAGCTACTTGGAGGCCAGGCAAGAAGATCGCTGAACTCGGGAGGCCAGGTTGCAGT
 GAGCTGAGAACATGCCACTGCACTCCAGCCTGGGACAAGAGCGAAACTCTGCTCTAAA

 7457 AGAACATCGCTGAACTCGGGAGGCCAGGTTGCACTGAGCTGAGAACATGCCACTGCACT
 CCAGCCTGGCAACAAGAGCGAAACTCTGTCTCAAGAAAATAATAATAATAAAAA
 ATAATAAAAGAGGGGGCATATGGGTGAAGTATGGACAAAATAGTGGGCAGGCACAGAT
 GATCTGGACACAGGAGCCCTGGAGTTATTCTGTAATCTAAGTGTTCATCTTATTAAA
 TATTGTGGCATAACACTCACAACAATAGCCAACACACCTCTTTGGAGCTTTATC
 [G, A]
 AAGTTTCCCACGTAAAGTTTTTCCCGCTTGTGATGCCGGTGGGTGGTGTGAA
 GCAGGCTTACGGGTGGCAGTTCTCACAAGGCATTAAGTGGCCTTGTCTAGGTCTGC
 CTTCAGCGAGGATGACGGTGAAGCAGGCCAGGCTTGGAGTTGTCAGGGCCACAAGGGGACAGGCTGGT
 CTCACAGCTCAAATGGTCTGGAGGATGAGAACAGGCAATTGGTGAGCAGACACCAC

 7830 GGTGGCAGTTCTCACAAAGGCATTAACGGCCTTGTCCAGGTCTGCCCTCAGCGAGGA
 TGACGGTGAACGCCAGGGCAAGCCTTCCATGGGCCAGCTGGCCCTCTACCTGCTCGCTCT
 CAGAGCCAACACTGAGTTGTCAGGGGCCACAAGGGGACAGGCTGGTCTCACAGCTCAA
 ATGGTTCCTGGAGGATGAGAACAGGCCATTGGTGAGCAGACACCACATCCGCTGGGGTGG
 GGAGCAGCTGGAGGCCATCAGATGATATTCTCAAATGAGAACATCAGAACATTGGGTTT
 [T, A]
 CTCCCCAGCGTCTTCCCACCATCCATTCTGCCATCTCACTGCTCTAGCTAGAGGCTCG
 AACCTGTCCTCCATGCCATCTTGACCCAGCTTCTCCCGCGCTGCACACATACTATTGAC
 AGGTGTGTTCTGCTGTTTTGTTGTTGTTGTTGAGTTGGAGGTTTG
 TCTTGCTGCCAGGCTGGAGTACAATGGCGCAATCTCAGCTACCGCAATCTGCCCT
 TGGCTCAAGCAATTCTCTGCCCTAGCCCTTGAGTAGCTGGATTACAGGCATGCC

 8089 ATCAGATGATATTCTCAAATGAGAACATCGAACATTGGGTTTCTCCCCAGGCCTTT
 CACCATCCATTCTGCCATCTCACTGCCCTACGTAGAGGCTCGAACCTGCCCCATAGCCA
 TCCCTGACCCAGTTTCCCGCGCTGCCACACATACTATTGACAGGTGTGTTCTGCTGGTTT
 TTTGTTTTGTTGTTGTTGTTGAGTTGGAGGTTGCTCTGCTGCCCTGGTTCAAGCAATTCTC
 [T, C]
 TGCCCTAGCCCTCTGAGTAGCTGGGATTACAGGCATGCCACCAACCCAGCTAATT
 GTATTTTAGTAGACGTGGGTTCTCCATGTTGGTCAGGCTGGTCTCGAACCTCTGAC
 TCAGGTGATCCGCTTGCCTAGCCTCGAAAGTGTGGGATTACAGGCATGCCACTGC
 GTTGGCCCAGTACAAGCCTGATTGGCTAGCCACCAAGATTGACTTGATTATCCACC
 TTGGGACAACGGACAGCCTGCTTATGACTTACGCCATAGTCTCTACTAGCTCTC

 8551 TACAGGCATGAGGCCACTGCGTTAGGCCACTGACAAGCCTGATTGGCTAGCCACCAAG
 ATTGACTTGTATTATCCACCTCGGGACAACCTGGACAGCCTGTTATGACTTACGCCATAG
 TCTGTCTCTACTAGCTCTGCCCTGACTTGACCCAGCATACACAGCCAGGCCAGCC
 TTTTCAATATAAACCTGATCTGCTGCCACTGCTAAACCTGCAGGGGCTGCCACTGC
 TCCATGGCCAGCCTGCTACCCCTACCTCTGCCAGGCTGCTCATCCATTCTCTGC
 [C, T]
 TCCCACACACCTGCCCTCTGGGCTCCAGCCATACCATCTCTCAACTCATAAGCCAGTT
 TTTTCATACAGGCTCCCTCATCTGGACTGGCTTCCCTGCGTGCAGTTCACTCTGCTCT
 ACCTTGGCTCTGCCACCCATCTCAGCGTCTCCAGCATTACCTCTGGAGAACATC
 CTGCTTGACTTCCCAGCCACCCAAATCACTACTTGGTCTGCAATTCTCGTGTGCAATTG
 CAGTCGCATGAGCAATTGCTGTGGTTGAGGCCGAACCGCAGTGCCTGCTGCCATG

FIGURE 3K

9269 AGGCCAGGGTCCCAGGTGCTGGCGGGCTGGCTGCTGGGGGGCAGAGAGGCCACCCCC
 TCTGTTTTTTCCTCTCAGGGCATGATCACAAAGGGCACCCCCCACTAGCTACTACCA
 GTATGGCCTGGCATCTGGCCCTGTCTCCACCAAGAAGCGGGTCCATGACAGCGTGGT
 GGACAAACTCTGTATGCTGGAACCTTCCACCAAGGGCCACCATTGTGGGTGAGTA
 GGTCAAGACCGTGCAAGGCCAGGCTGGCACTCCCTCAGTCCCAGGTCTGCACGTGATGAC
 [G, C]
 TCCATACCTGGCCCCACACTCACCTTCCTGGGCTCTCGAACATCAAGTCCTTAG
 GGACGAATTGGCGAGGGCTATGGGTGATGCTCCAGCTGTGAGCCAGCTGGAGCTGGT
 AGGTGGATCTTGAGGCCAGGAGTTCAAGACAACGTGTGAAACCCCCATCTACTAAA
 AAAAAAAAAGTTAGCGGGCATGGTGCACATGCCTGAGTCCCAGCTACTCGGGAGGCT
 GAGGCAGGAGAATCACTGAACCTGGGAGGCGGAGGCTCAGTGAGTGGAGATCGCACCA

 9362 GGGCCACCCCCACACTAGCTACTACAGTATGGCCTGGCATTCTGGCCCTGTGCTCCA
 CCAGAAGCGGGCTTGACAGCGTGGACAAACTCTGTATGCTGTGAAACCTTCCA
 CCAGGGCCACCAATTCTGTGGGTGAGTGTGAGCCAGCTGGCAAGGGCAGGCTGGCACTCC
 CTCAGTCCCCAGGTGCACTGATGACCTCATAACCTGGCCCCACACTCACCTTCC
 TGGGGCTCCTCGAACATCAAGTCCTTAGGGAGCAATTGGGAGGCTATGGTGTGATGCT
 [C, T]
 CAGCTGTGAGGCCAGTTGGAGCTGGTAGGTGGATCTCTTGAGGCCAGGAGTTCAAGACA
 ACCTGGTGAAACCCCCTACTAAATAAAAAGTTAGCCGGCATGGTGCACATG
 CCTCTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAACTTGAACCTGGGAGGCGG
 AGGCTGCAGTGAGTGGAGATTCGCAACTGCCCTCCAGCTGGCAACAGAGTGAAGTGAG
 ACTCTGTCTAAAAAATAAAAATAAAATACCTCCCTAGTGATTCCAATGTGCAGCT

 9782 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAACACTTGAAACCTGGGAGGCG
 GAGGCTGAGTGAGTGGAGATCGCACCACTGCCCTCAGGCCAGGGCAACAGAGTGAGTGA
 GACTCTGTCTCAAAAATAAAAATAAAACTCCCTAGTGATTCCAATGTGCAGC
 TAAGTTGGAAATAGTGTTGATGGGTCAAGTCCTCTGGGCTCCCTCCAGTCCT
 CTCCCTAACCTCTAGCCCTCAAGTGCAGAGTGATCACGCCAACAGTAGTTGCCAGAAAAT
 [G, T]
 AGCAGTTTCTGGGACACAGGATTTAGAGTCCAGACAAGGAAAGTCTTGGGAGACCA
 GGTTGAGTTGGTGCCTTAGCTGATCTGACCATGTTGCCCTCTCTCCAAGCCCTCTG
 TGGTTGTCATAGCTACAAGGGCCTGACCCCTCAAGCCCTGCCCTGCTGGGCTTGG
 CTCTCCAGCTCATGTCATGTTCTGCCCCACTTCAAGACACAGCAGCATGGCAGGCTT
 GCATTCACCTGTCAGCGCTAAACCTCAACCTGGCGAGACAACGGATCACCAT

 11493 AAAAAAAAAATGGAGAAGAAGGAAGCTGGACATGGGCTCGTGTATAATCCTAGCA
 CTCTGGGAAGCTGAGGCAGATGGATTGCTGAGGCCAGGAGTTGAGACAGCTGGGCA
 ACATGGTGAAACCCCTGTCTTACTAAAATACGAAAGATTAGCCAGGCATGGTAGAC
 CCTATAATCCCAGCTACTAGGGAGGCTGAGCCACAAGAACACTTGAAACCTGGGAGACAG
 AGGTTGCACTGAGGCCAGATCGCCTGACCTCCAGCTGGGAGCAGTGTGAGACTC
 [G, A, T]
 GTCTCCAGAAAAAAACAAGAATGGATAGAGTGGAGCCAAGAAGAGGCAGGAAGAACAAAGA
 CACAGAGGTGCACAGAGTTGGGAAATTGGAGGAATGGCTTGCAGGAGATGGGATC
 TGGGAGAATGAGTGGAGTGGAAAGCAGATGAATGAAGAGAAGGTGAGCGCATCAGGTA
 ACAGAGATGGCTGTGAACAAATGCATGTTCTAGGAAGAGCCCTCTGGAGTGCTAGGTG
 CAGAGAGGTGGAGGAAGGAACTGGAGAGAACAGTGAGGGCCTGATCTTGG

 12260 ACAAGAAGACAAATAATCCAGGCTCTGCTCTCACACCAAGCTGCCGCCCTTCTTCC
 TGGCACAGTCATGTTGGAACAGCAGCTGCTGAGACCATCTCTCAGACCCAAGAGATCATCAG
 TGTCAGCTGCAAGGTCTTAGTCTCTGCGCCGTACAGACAGTCATCTGTCTGGC
 CGGGTCCACCGTGGAAAGATGCTCTGAAAGAAGGCCCATGAGTTAGGAGGATTCAAGTGA
 CTCCACCTCCAGTCTCACCCACCAACCTCACATGCCGATAACAGGGTCAACAGAA
 [A, G]
 AGACGGGAAACAGAGGAGAGGGTCCCTGGGAGAGACACTGCCCTGCTCTGTTCTA
 CCTGCTCAGCTCCTTCTTGCCCACGGTGTATGGAAACAGGGAGCCATAGGCCAGCATT
 GTCACTGAGAGAGCAGGCTTGGAGGCGAGGCCCCCAGTGGAAATCCAACCTAACC
 GCTAGTTCCAGGTAGCCACCCACAATTACCGAGGAGAACAGTTGTGCCCTCCCTGC
 AGGGCCAGTGTGAAGAGTCAGGAGTTAGTACACATAGAGATAGTGGCATGTGCTTTTA

 13086 GGCACGTGCCACACAACCTGGTAATTTTTTTTTTGAGATAGGGCTCTG
 TCTGTTGCCAGGCTGCTCAAATTCTGGCCTCAAACCATCCTCACACCTGAGGCGCT
 CAAATAATTGGGATTATAGTGTGCGAGGCCATGCTCAGCCAGAATAATACTGGTTTT
 TTTGTTTTTTTTGAGACAGAGTCTCACTCTATTACCCAGGCTCTGGAGGCCAACCTCG
 TGTTGTGTTAGTTGTTATTATTTATTATTTATTTCGAGACAGGCCCTCTCTTT
 [T, C]
 ACCTAGGCTGGAGTGCAGTGGCGCAATCGGCTCACTGCAACCTCCGTCTGGGTT
 AAGTGATTGTCCTGCCCTGCCAGCCTGAGTAGCTGGCTCACAGGCCGTGCCACCATGC
 CCAGCTAATTGGTATTAGTAGAGACAGGGTTACTATGTGGCCAGCTGGTTTC

FIGURE 3L

TAACTCCTGAACCTGGGTGATCTGCCTGCCCGCCAAAGTGCCTGGGATTACAGGC
 ATGGGCCTCCGTGCCCGCCATGTATTTAGGCAAGGTCTCTCTGTATCCAGGC

13183 ACCATCCTCACACCTGAGGGCGCTAAAATATTGGGATTATAGGTGCGAGGCATCATGCTC
 AGCCAGAATAAAACTGGTTTTTTGTGTTTTTGAGACAGAGTCACCTCTATTAC
 CCAGGCTCTGGAGGCCAACCTCGTGTGTTGTATTTTATTTTATTTATTTATTT
 TTGAGACAGAGCCTCTCTTACCTAGGCTGGAGTGCAGTGGCAATCTCGGCTCA
 CTGCAACCTCCGTCTCTGGGTCAGTGAATTGCTGCCTCAGCCTCTGAGTAGCTGG
 [T, C]
 GCTACAGGCGCGTGCACCAGGCCAGCTAATTGGTATTTAGTAGAGACAGGGTTT
 TACTATGTTGCCAGCTGGTTCTAACCTCTGAACCTGGGTGATCTGCCTGCC
 CCCAAAGTGCCTGGGATTACAGGCATGGGCCCTCGTGGCCAGCTAGTATTTAGGCA
 AGGTCTCTCTGTATCCAGGCTGAAGTGCAGTGGCACATTCACTGCAGCCT
 CAAATTATCCAAGTAACAGGGACTACAGGCATGCACCACACCCATCTACTTTTTT

21240 TCAGCTCACTGCAACCTCTGCCCTGGCTTAAGGCATCTCCCACCTCAGCCCTCTGA
 GTACGTGTGACCATAGGCCATGGCACAAGGCCAGCTAATTGGTATTTAGTAGA
 AATGTGGTTACCATGTTGCATAGGCTGGTCTCGAATTCTGAACATAGTGAATCTGCC
 TGCCCTGGCCCTCCAAAGTGCCTGGGATTCTAGGTATGGTACGCCACCCGCTGGCTATAAT
 GGCACTTCCATCCCATTGATGAGGCTACTCTCATGACCTAATCATCTCCAAAGGC
 [C, G]
 CTAAGGCCTCTGTACCATCACCTGGGGTAGGTTAACATATACATTGGGGGG
 ACACAGACATTAGACCATAGCACCTCCATTGAAAGGAAACATTCTGACACCTGGCTA
 TCTCAAAGGGCCCTTCAGTCCCTGCAGGCTGCATTCCACATACCAACAAGAGCAG
 CGACACTCACTCAGAGGTTAATAACTTGTCCAGAGTCACAGCAGTAATGAATGACAGAG
 CTGGGGCTGAATCCAGGCGCTCTCTAGAGCCTGGATTCTGTAGTAGTGAAGCTG

21695 CATTCCCACATCACCAACAAGAGCAGCAGACTCACTCAGAGGTTAACATTGGTCCAG
 AGTCACAGCAGTAATGAATGACAGAGCTGGGCTTGAAATCCAGGCCCTAGAGCCT
 GGATTCTGTGAGTGAGTGAAGGCTGACTCTGGGAGACTCTGGTGGTCTGGTCTC
 TCTCCAGACTGCAGTGCAGTCTCTGGTCTAGGGTCCCTAGGGTATTACAAAGACA
 GTGCCCTGCCTGTCAGGTGTTTATACAGATGAGGTATGGCTAGGAAACCTGT
 [A, G]
 GGAAGCTGAGTTAGACTTTGAGCAGGCTTAGGGAGGTTCCAGCTCCACCAAA
 GCCCCAGGTGATTCTACAGACTCTAGCCTCAGGGTGGGGCTGGAGATGAGGTTG
 CGGGGTGCGATATTCTGCCAATTGCCCTCTGGTCAATCTGTTCTGAGGTATTG
 CTGACTACAGACCCAAGGATGGAGAAACATTGAGCTGAGGCTGGTAGCTGGTAGCCCC
 TGAGCTCCCTCATCCCAGCAGCCTCGCACACTCCCTAGGTTCTACCCCTCCTGATG

22058 CCCAGGTGGATTCTACAGACTCTAGCTCAGGGTGGGGGCTGGAGATGAGGTTGCG
 GGGTGCATATTCTGCCAATTGCCCTCTGGTCTGCAATCTGTTCTGAGGTATTGCT
 GACTACAGACCCAAGGATGGAGAAACATTGAGCTGAGGCTGGTAGCTGGTAGCCCTG
 AGCTCCCTACTCCAGCAGCCTCGCACACTCCCTAGGTTCTACCCCTCCCTGATGTC
 CCTGGAACAGGAACCTGGCTGACCCCTGCTGCCACCTCTGTGCACTTGTAGCAATGCC
 [C, T]
 TGGGATCACCCAGGACAAGCCCTCGAGGGCCCTATACCATGGCCACCTTGGAGCAG
 AGAGCCAAGCATCTCCCTGGAGTCTTCTGCCAAGTCTGGCCAGCCTGGCCCTGCA
 GGTCTCCATGAAGGCCACCCATGGTCTGATGGCATGAAGCATCTCAGACTCTTGGC
 AAAAACGGAGTCAGGCCAGGGCGCAGGTGTTGAAGACCAACTCGTTCTGTGGGGTC
 CTGCAAGAAGGCCCTCTGAGCCGGGGCTATGCCCTGACCCAGCTCCACTCTGCT

22233 CCCTGAGCTCCCTCATCCCAGCAGCCTCGCACACTCCCTAGGTTCTACCCCTCCCTG
 ATGTCCCTGGAACAGGAACCTCGCTGACCCCTGCTGCCACCTCTGTGACTTTGAGCAAT
 GCCCCCTGGGATCACCCAGGCCACAAGCCCTCGAGGGCCCTATACCATGGCCACCTTG
 GAGCAGAGGCCAGCATCTCCCTGGAGTCTTCTGCCAAGTCTGGCCAGCCTGGC
 CCTGAGGTCTCCCATGAAGGCCACCCATGGTCTGATGGCATGAAGCATCTCAGACTC
 [C, A]
 TTGGCAAAAACGGAGTCAGGCCAGGTGTTGAAGACCACTCGTTCTGTGGTTG
 GGGTCTGCAAGAAGGCCCTCTGCCAGGGGGCTATGCCCTGACCCAGCTCCACT
 CTGCTGTTAGAGTGGCAGCTCCAGCTGGTAGCTGGCAACTAGCTGGGGAGACCTCAGCA
 GGGCTGCTCAGTGCCTGCCCTGACAAAATTAAAGCATTGATGCCCTGTTGGACCTGCTAC
 AGTGGCCTGGTGCCTCATACTCCTCAGGTGCAAGGGCAGGGACAAGAGAAGGGGAAGTA

22245 TCATCCCAGCAGCCTCGCACACTCCCTAGGTTCTACCCCTCCCTGATGTCCCTGGAA
 CAGGAACCTGCCCTGACCCCTGCTGCCACCTCTGTGCACTTTGAGCAATGCCCTGGGAT
 CACCCAGCCACAAGCCCTCGAGGGCCCTATACCATGGCCACCTTGGAGCAGAGGCC
 AACGATCTCCCTGGGAAGTCTTCTGCCAAGTCTGGCCAGCCTGGCCCTGCAAGGTCTC
 CCATGAAGGCCACCCATGGTCTGATGGCATGAAGCATCTCAGACTCCTGGAAAAAA
 [C, -]

FIGURE 3M

GGAGTCCGCAGGCCGAGGTGTTGTGAAGACCCTCGTTCTGTGGTGGGTCTGCAAG
 AAGGCCTCCTCAGCCCCGGGCTATGCCCTGACCCAGCTCTCACTCTGCTGTTAGAG
 TGGCAGCTCCGAGCTGGTGTGGCACACTAGCTGGGAGACCTCAGCAGGCTGCTCAGT
 GCCTGCCTGACAAAATTAAAGCATTGATGGCCTGTGGACCTGCTACAGTGGCTGGT
 CCTCATACTCCTCAGGTGCAGGGCAGGGACAAGAGAAGGGGAAGTAACCCATCAGGG

 22375 ACAAGCCCTTCGAGGCCCTATACCATGCCAACCTGGAGCAGAGAGCCAAGCATTTC
 CCTGGGAAGTCTTCTGGCCAAGTCTGGCAGGCCCTGGCAGGCTCAGGCTCCCATGAAGGC
 CACCCCATGGTCTGTGGCATGAAGCATCTCAGACTCCTGGAAAAACGGAGTCCGC
 AGGCGCAGGTGTTGAAAGACCACTCCTCTGTGGTGGGTCTGCAAGAAGGCCCTCC
 TCAGCCCAGGGCATGGCCTGACCCAGCTCTCACTCTGCTGTTAGAGTGGCAGCTC
 [C, T]
 GAGCTGGTTGTCAGTAGCTGGGAGACCTCAGCAGGCTGCTCAGTGCCTGCCTCT
 GACAAAATTAAAGCATTGATGGCCTGTGGACCTGCTACAGTGGCCTGGTCCTCATACTC
 CTCAGGTGCAGGGCAGGGACAAGAGAAGGGGAAGTAACCCCATCAGGGAGGAGTGGAG
 GTGCGCTGAGGCCATGGGGAGTGGAAATGCCAGCAGTGTGACGT
 TGACTACTGACTGAGCACCCACTACTATGACTGAGCACTCACTGCTAGATACTATCTTG

 23042 GCCGGCGTGGCTCACGCCCTGAAATCCCAGCACTTATTAGGAGGCCAAGGCAGGTGGAT
 CACAAGGTCAAGGAGTTGAGATCAGCCTGGCAACATGGTAACTCCATCTTACTAAA
 AATACAAAAAAATTAGCCAGGCATGGCTTGTGCATGCCATGCCATGTAATCCCAGTTACT
 TGGGAAGCTGAGGAGGGAAATTGCTGAAACCTGGAGGGAGGTTGAGTGGCAGGCCAG
 ATCACGCCATTGCACTCCAGCTGGCAAGAAGAGAAACACTCTCAAAAAAAAAAAAA
 [A, -, T]
 CAGGAAACTGGTGCCTCAAAAGGAAAAGTGACTCACCAGGTACAGACTAGGCAGTGAT
 GCTGGGGAAACCTGGCTCAGGGACACAGACACTGGCCTGGGCAGCCTTGAGCCTCCTCC
 ACTAAAATACTGAAAATGAGGGGCTTCGATGATGGTTATAATCGTATGGCAGAGCCCCAA
 CTCAACTGGAGCCCTGGACCCAGAACGCTAGGGTCTCACTCCCTGCTTTCCACAAAGGCA
 CCATTAGGGCATCCCCAGGCCCTGGCAGCCACGACGAGGATCCTGCCCTCATGGT

 23344 AGGAAACTGGTGCCTCAAAAGGAAAAGTGACTCACCAGGTACAGACTAGGCAGTGATG
 CTGGGGAAACCTGGCTCAGGGACACAGACACTGGCCTGGGCAGCCTTGAGCCTCCTCCA
 CTAAAATACTGAAAATGAGGGGCTTCGATGATGGTTATAATCGTATGGCAGAGCCCCAA
 TCAACTGGAGCCCTGGACCCAGAACGCTAGGGTCTCACTCCCTGCTTTCCACAAAGGCAC
 CATTAGGGCATCCCCAGGCCCTGGCAGCCACGACGAGGATCCTGCCCTCATGGT
 [T, C]
 GGGGGCTTAGGGCTCTGGCTGCCCTTGAAGAGGGGTTAGGCCAGCGAGGCACCC
 CCTATGCTGACCCCCAACAGGTTAGGAAGAGGTCTGTCTCAGTGGGCCCTGTGATG
 AACAGCCCACAGGTCTGCTCCACATGCCCTGGAAAGAGATGGTGAACATACTCAAAGTCC
 TTGAAGCCGATATTAAACCCACCTAGAGCACCATTCTCAAACATTAGGGTCTGAGAAGA
 TAGGGAAAGTAAGCAATTAAAACATTCTTATATTGGCCAGGTGCAATGGCTCACGT

 23873 GGTCTGAGAAGATAGGGAAAGTAAGCAATTAAAACATTCTTATATTGGCCAGGTGC
 AATGGCTCACGTCTGAAATCCCAGCCTTGGAGGACGAGGATCACCTGAGGTAGGAG
 TTCAAGATCAGCCTGGCAACATGGAGAACCCCATCTCTACTAAAATACAAAATTAG
 CTCAGCGTGGTGTGACCTGTAATCTAGCTATTAGGAGGCTGAGGCACAGAACAT
 TGCTTGAGTCATATTGACCACTGCACTCCAGCCTGGCAACAGCGAGACTCTGTCTC
 [A, -]
 AAAAAGGAGATATTGCTGAAAAGACCCAGCCTGCCAAACTCAGGGCAGCCAAGG
 GAGGTAGTGAATGGAGTGGAGCTCAGCGCTCCACCTCCACTGCCCTCAGGCCCT
 CTCTGCCCTTCCCATCAGCTGCTTCTGGCATGGCCTGGCAGAGACTTGGCCT
 CCTCCAGTCAAGCTCCTCTTAGATTGCTCCACGCCAGTGAAGTCTTGGGACACTG
 GGTAGATGCTAGTCTGGCACAATTGGCAGGAATCCAAGAACAGTGTGAGTGGGG

 24764 ATAGCTGGATTACAGGTGTCACCACCATGCCAGCCTAATTGGTATTATTAGTAGA
 GATGGGTTTCACCATGTTGTCCAGGCTGGCATGAACCTGACCTCAAGTGATCCACC
 CGCTTGGCTCCCAAGTGTGGATTACAAGCATGAGCCACAGTGCCTGGCCTGACCC
 TGCTCTTGAAGACCATCCCCAAATTCTGTGCACTGTGCTTCTCCCTCTGCCAAATCCCAC
 CCTCTCTCAGCTCTGCCCGCTCTCCCTCTGCCAAATCCCACATCTCT
 [G, T]
 GAAGCCCTTCTCCAGGGAAAGCCCTGATCATGCTGTTCTCCTGTGGAGGGATGAAG
 GACGTGCCACGGAGTTGTTGTTGAGATGGAGTTGCTCATGGTGCCT
 AGGCTGGGTACAATGTCAGCTCAGCTACTGCAACCTCTACGTCCGGTCAAGC
 GGTCTCTGCCCTAGCCTCCAGTAGCTGGGATTACTGGCATGAACCACACACTGG
 CTAAATTGTTAGTAGAGATGGGTTCTCATGTTGGTCAAGGCTGGTCAAC

 24939 GACCCCTGCTTTGAAAGACCACTCCCCAAATTCTGTGCACCTGTGTCCTTCT
 CTCTGCCCTCTCAGCTCTGCCCGCTCTCCCTCTGCCAAATCCCAC
 TCTCTTGAAGCCCTTCTCCAGGGAAAGCCCTGATCATGCTGCTTCTCCCTGTGGGAGGG

FIGURE 3N

ATGAAGGACGTGGCCCACGGAGTTGTTTGTGTTGAGATGGAGTTGCTCATG
 TTCCCCAGGCCTGGGTACAATGGTACGATCTCAGCTACTGCAACCTCTACGTCCGGGT
 [T, C]
 CAAGCGGTTCTCCCTGCCCTAGCCTCCCCAGTAGCTGGGATTACTGGCATGAACCACCA
 CCTGGCTAATTGTGTTTAGAGAGATGGGTTCTCATGTTGGTCAGGCTGGTCT
 CGAACCTCCAACCTCAGGTATCTGCCCTGCCCTCCAAAGTACTGGGATTACAGG
 GTTGAGGCCACTGCCCTGGCCACGGAGTTAAGAGGCTTCTGTGGCAGTGG
 CATCCAGACGGAGTGCAGAAACTCAAAGTTGAAGGCCAGAAGCTCAGGGAAAGGGGAGTG

 24945 GCTCTTTGAAAGACCATCCCCAAATTCTGTGCACCTGTGCTGCTTCTCTCTGC
 CTCTCTCAGCTGCCCTGCCCTCTCCTCTGGCAAATCCACTCATCTCT
 GAAGCCCTTCTCCAGGGAAAGCCCTGATCATGCTGCTTCTCTGTGGGAGGGATGAAG
 GACGTGGCCACGGAGTTGTTGTTGAGATGGAGTTGCTCATGTTGCC
 AGGCTGGGTACAATGGTACGATCTCAGCTACTGCAACCTCTACGTCCGGTCAAGC
 [G, A]
 GTTCTCTGCCCTAGCCTCCCCAGTAGCTGGGATTACTGGCATGAACCACACCTGGC
 TAATTGTGTTTAGAGAGATGGGTTCTCATGTTGGTCAGGCTCGAAGT
 CCCAACCTCAGGTATCTGCCCTGCCCTCCAAAGTACTGGGATTACAGGGTTGAG
 CCACTGTGCCCTGGCCACGGCCACGGAGTTAAGAGGCTTCTGTGGCAGTGGCATCCA
 GACGGAGTGCAGAAACTCAAAGTTGAAGGCCAGAAGCTCAGGGAAAGGGGAGTGTGAGTT

 25092 ATCATGCTGCTTCTCCTGTGGGAGGGATGAAGGACGTGGCCACGGAGTTGTTGTTGTT
 TTGTTTGAGATGGAGTTGCTCATGTTGCCAGGTGGGATCAAATGGTACGATCTCA
 GCTCACTGCAACCTCTACGTCGGGTTCAAGCGGTTCTCCTGCCCTAGCCTCCCCAGTA
 GCTGGGATTACTGGCATGAACCACACCTGGCTAATTGTGTTTAGAGAGATGG
 GGTTCTCATGTTGGTCAGGCTGGTCTCGAACCTCAGGTATCTGCCCT
 [C, T]
 GGCCCTCCAAAGTACTGGGATTACAGGGTTGAGCCACTGTGCCCTGGCCACGGGA
 GTTTAAGAGGCTTCTGTGGCAGTGGCATCCAGACGGAGTCAGAAAACCTCAAAGTTGAA
 GGCCAGAAGCTCAGGGAAAGGGGAGTGTGAGTTGAGGACTCTTGGCTGCCAGGGCAG
 AAACCGAACCTCAAGCTCTCCACAACAGGGGTGTAGACATGTAGAATCAGAGAGGAG
 GCTGAGGCCATGCAGCCCCGAGAAGAGGGAAATGCCACTGAGCCACAGAGACCCAGTGCCA

 25428 AGTGCAGAAACTCAAAGTTGAAGGCCAGAAGCTCAGGGAAAGGGGAGTGTGAGTTGAGGA
 GTCTCTGGCTGCCAGGGCAGAAACCGAACCTCAAGCTCTCCACAACAGCGGGTAG
 ACCATGTAGAATCAGAGAGGAGGTGACCCATGCCAGGGCTGAGGAAAGAGGGAAATGCCACT
 GAGCCACAGAGACCCAGTGCCACTGCCAGGTGTCTCTGCCCTCACCTCCCATGACCC
 [T, G]
 GCCTGTCTGTATGAGGCTTCACCCCTCTCGTTGACATTGACACATTCTAGGTGA
 CACCAGCAGCTCTGATTCTCATCTCCATAACATCAGCCCCCAGAGAGGGACAAC
 CTGAGCTGATAACATAATAGATGCCCTTCTGGAGGGCATGGTCATGGTCAGCGTGG
 GAGGATGAAGCTGAGCAGGCAGGATGGGGTCTAGAGGGAAAGGAGGTGAAAGTT

 25513 GCCAGAAGCTCAGGGAAAGGGGAGTGTGAGTTGAGGAGTCTTGGCTGCCAGGGCAG
 AAACCGAACCTCAAGCTCTCCACAACAGGGGTGTAGACATGTAGAATCAGAGAGGAG
 GCTGAGGCCATGCAGCCCCGAGAAGAGGGAAATGCCACTGAGCCACAGAGACCCAGTGCC
 CTGCCAGGTGTCTCTGCCCTACCTCCATGCCAGGGCTGTCCTGTATGCAGGCTTCA
 CCCTCTCTCGTTGACATTGACACATTCTAGGTGACACCAGCAGCTCTGATTCTCATC
 [C, T]
 CCCATAACATCAGCCCCCAGAGAGGGACAACCTGCTGAGCTGATAACATAATAGATGCC
 CCTTCTGGAGGCCATGGTCATGGTCAGCGTGGAGGGATGAAGCCTGAGCAGGCAGGA
 TCAGGGGTCTAGAGGGAAAGGAGGTGAAAGTTGAGATCACAGACCTGTGGTCAGGTGG
 TGGGAAGGGTTGACGAGTGTGGCCCAAAGAGCTTGGAAAGGGATTGCTGCTGTGG
 GAGCACTGCCCTCCCTAGGGACAACAGCCACCTTCTCTCCCCATTGCTCTTCCC

 25684 CAGGTGCCACTGCCAGGTGTCCTGCCCTCACTTCCCATGACCCGGCTGTCTGTATG
 CAGGCTTCACCCCTCTCGTTGACATTGACACATTCTAGGTGACACCAGCAGCTCTG
 ATTCTCATCTCCATAACATCAGCCCCCAGAGAGGGACAACCTGCTGAGCTGATAACAT
 AATAGATGCCCTTCTGGAGGCCATGGTCATGGTCAGCGTGGAGGGATGAAGCCTGA
 GCAGGCAGGATGGGGTCTAGAGGGAAAGGAGGTGGAACTTGAGATCACAGACCTGTGG
 [C, T]
 CAGGTGGCTGGGAAGGGTTGACGAGTGTGGCCCAAAGAGCTTGGAAAGGGATTGCT
 GCTGAGGAGCAGCAGCTGCCCTCCCTAGGGACAACAGCCACCTCTCTCTCCCCATT
 GCCTTCTCCCTCTGTAGATATGAAACACAGGCCCTCTGTGAGGCCCTACTTAACCTCC
 GTGATGGGAAAGCGGCCAGAAAGGGAGTTCTGGCAGCTCTCCGAGACCCAAACACC
 CCACTGTTGCAAGGTGAGTCATGCCCTGACACTCTGGATGTGCCCCCTACCCAGCTTA

 26165 GTGATGGGAAAGCGGCCAGAAAGGGAGTTCTGGCAGCTCTCCGAGACCCAAACACC
 CCACTGTTGCAAGGTGAGTCATGCCCTGACACTCTGGATGTGCCCCCTACCCAGCTTA

FIGURE 30

CTCAGCCAAGAGGCTTCATCAACTCACCCCAGCTTCCCTAGCACCCCTGGGCCACAC
CTTCACAAAATCACTGATGCTAAAGTGGATATAATATATTGAACGTAGCCTAGCAT
TTTATGCAAGTTACTGTGAAATTCTAGGAAACCAGACAGATTACAAAAAAAAAAAAA
[A, -]
CTAGAAGAAAATTAACATCACCTAGGATATACTACCTAGGAATAACGTCTTTATTTGA
GATGGAGTTCGCTCTGTTGCCAGGCTGGAGTGCAACGGTATGATCTGGCTCGCTGC
AACCTCCGCCCTCTGGGTTCATGTGATTCTCCACCTCGGCCTTCAGAGGCCAAGTGG
TCTGCCTGCCCTGCCTCCAAAGTTGGGATTACAGGCATGAGCCACCGCACCCAGCC
AAAATTACTTAACCTTTCTAGATACTTTTAAAAATATGGCAGTAAGTTTCATAAA

FIGURE 3P

Isoform 2:

FEATURES:

Exon: 2132-2195
 Intron: 2196-5670
 Exon: 5671-5863
 Intron: 5864-7672
 Exon: 7673-7761
 Intron: 7762-9149
 Exon: 9150-9302
 Intron: 9303-10101
 Exon: 10102-10274
 Intron: 10275-10399
 Exon: 10400-10586
 Intron: 10587-12128
 Exon: 12129-12294
 Intron: 12295-25922
 Exon: 25923-26040

Allelic Variants (SNPs):

Position	DNA			Protein		
	Major	Minor	Domain	Position	Major	Minor
1022	C	T	Beyond ORF(5')			
1882	C	T	Beyond ORF(5')			
1951	G	A	Beyond ORF(5')			
2940	A	G	Intron			
3831	G	A	Intron			
6732	G	A	Intron			
7558	G	A	Intron			
7931	T	A	Intron			
8190	T	C	Intron			
8652	C	T	Intron			
9370	G	C	Intron			
9463	C	T	Intron			
9883	G	T	Intron			
11594	G	A T	Intron			
12361	A	G	Intron			
13187	T	C	Intron			
13284	T	C	Intron			
21341	C	G	Intron			
21796	A	G	Intron			
22159	C	T	Intron			
22334	C	A	Intron			
22346	C	-	Intron			
22476	C	T	Intron			
23143	A	- T	Intron			
23445	T	C	Intron			
23974	A	-	Intron			
24865	G	T	Intron			
25040	T	C	Intron			
25046	G	A	Intron			
25193	C	T	Intron			
25529	T	G	Intron			
25614	C	T	Intron			
25785	C	T	Intron			
26266	A	-	Beyond ORF(3')			

Context:

DNA
 Position

1022 TTGGAGATATTAAAGGTCATAGTGTCTCACAAATTGAGCTGAAAGGAACTGTTAGGA
 TGATCTTGCTAACCCCTCATCTCACACAGGAAGAACTATTTAAACTCGAGAGGTTAA
 GTGACCTGGCAAAGTCACACAGCCACCACTAGTTAACCTCGTATACTGATTCTCCTGT
 GGGGCTGGGCAGATGAGGAATCTTTGTTCTCTCCCTGTTGCAGAGATTTTTTGAG
 GTTACTTCCGAGTTCTGGCAAGTACCCCTGCTTCTGGTAGCTTGTTGTCTCGATTCAAT

FIGURE 3Q

[C, T]
TCATTCTTTTATTTATTTGAGACAGGGTCTCACTTGTCAACCAAGCTGGA
GTGCACTGGGTAACTCTGGCTCACTGTAGCCTCCACCTCTGGGTTCAAGCGATCCCTCC
TGCCCTCAGCCCCCAAGTAGCTGGGATTACAGACGTCTGCCACCACGCCAGGCTAATTTA
TGGTTTTTGTATGTGTTTTGTGTTAGAGACAGTGTCCCCATGTTGCCAG
GCTGGTCTCAACTCCTGAGCTCAAGTGTACTGCCGCCAGCCTTCAAAGTGTAGG

1882 ACAGCTGACTCCAGCAATGCTGCTCACGTGACCCTGCAGCTGCAAGCTCCGTTCCACTC
CTTGCTCTGGCTAGGTGGCACTACCAAGGGCTCCTGGTAAGGAGTACGGGTAGGC
ACCCGGTCTGCCAATCCACCACTGGAACAGCTGGGGGACAGCAGACAGGCACGGTCGG
ACAGACTTGACAGATCAGGCATCAGGCCCTCTGCGCTGGTCCCAGGCTTTAACGAGGA
ACGTGAATGGCCTCAAGATGTCACATGGTCCCAGTGCCTCCCTCCCTTGTTC
[C, T]
TACCTCCAGGAGGGCTGCTCTGCCCTCCTCTGGCTTATGTTCCCGC
CACACAGGCCCTCCCCGCCACCCCTCTGCAGACTTAGCGTGCATTGCAGGCATGG
AGGATAATCAGTGACAGGAAGCTGCGCTCTCGAGGGTACCCAGCTGCTCAGGAG
AGCTCAGCAGGGCCAGCCCAGGAGTCTTCCGATTCTGCTCACTGTCACCCACCT
GCTGCTGCCATGAGGCACCTGGGCCCTCCTTCTGGGGCTTGTCCCTGGGGCTGGGCCCTC

1951 GGCTAGGTGGCACTACCAGGGCTCCTTGGTAAGGAGTACCGGGTAGGCACCCGGTCC
TGCCAATCCACCACTGGAACAGCTGGGGGACAGCAGACAGGCACGGTCGGACAGACTTG
ACAGATCAGGCATCAGGCCCTCTGCAGCTGGTCCGGCTCTTAAGCAGGAACGTGAATG
GCCCTCAAGATGTCACATGGTCCCAGTGCCTCCCTTGTCCCTACCTCCA
GGAGGGCTGCTCTGCCCTCCTCTGGCTTATGTTGGCTTATGTTCCCGCACCACAG
[G, A]
CCTTCCCCGCCACCCCTCTGCAGACTTAGCGTGCATTGCAGGCATGGAGGATTAAT
CACTGACAGGAAGCTGCGCTCTCGAGGGTACCCAGCTGTGGTCAGGAGAGCCTCAGC
AGGGCCAGCCCAGGAGTCTTCCGATCTTGCTCACTGCTCACCCACCTGCTGCTGCC
ATGAGGCACCTGGGCCCTCCTCTGGGGCTCTGGGGCCCTACTGAGATG
TGTGGTAGTAACCTGCCCTATCTGTGCCCTTCTGGTCTGGTCTAGTGGGTGG

2940 AACATAGGGAGACCCCATCTCTACAAAAAAATAAAAAAATAGCTGGCATGG
GGAAGACTTCTGAAGACCAAGAGGACACATGGGAGCTAAACTCGAAGGAAGAAAAGGA
GCTGGCAGGAAAGGAGTGGGGACACACATTCTAGGCAGCAGGAAGTGAACGGCTCGGAGG
TCTCTGCTCTCCAGCTGTGCCCCAAGGGGCTCTTGAGCAGTCAGTCTCTGGACCT
GTCTATGAGTCTGAGCTAGAGGTCAGGGCTGCTCCTCAGACAGGAGGAGAACAG
[A, G]
CTTGGGAACCTTGGGCCGCCACGCCCTTTCTCTCTCTGCACCTAGGATTACGTT
GAGCAATACACTTCACCCCCATGGTCTCTTGAGACCCCTGGGAAACCTCTGGAGGGTGG
TGCAGTCATGTCAGGTGCAAGTGAAGAAGTCGAGGGTGGAGGGCTGAGTGAACCCAC
TCAGGGTGTCCACCTTCCAGAGCTTGCTGAACCTAGTTAGAAACTTGAAGCCTC
GTTTGTGTTGTTGTTGTTGAGAGAGGTTCTCCCTCTGTTGCCAGGCTGGAGT

3831 GACACCTCAGGTCTGGGCCAGGAACCCAGCTCTGGTTCATGTCGGACAGTCCCCAG
GGGAGTTCTGGGTTCAACCAAGCAAGCTCTCCTCTGGCTGATCTGGTCTCAGCCTT
GGACAGTTAGTCATTAACCTGACCCACAGGAGCCCAATCCCTGGGCTGGGAAT
CTTGAACCTGGGTTGGGTGCAAATATCTGCACTGAGTCAGTAACTATTGCAACCAAGCCTC
ATTCTTATCTGAAAGTGGCTAAGAATGCTCCCTGCCCTCCTCGGTGAGTAC
[G, A]
AGGAAGGATCCCAGACACCTGCTCTCCAGTTAAAGCTTATGTTGAAAT
TGACAGGGATCGCTGCACAAACGCTAATGCAAAGTGGCTCTGTGCTTCTCTTCTCTT
CTCTCTCTTTTTTTAATTCTCTAGAGATGAGGTCCTACTATATTGCCA
GGTTGGTTCAAACTCCTAGGTCAAGCGATCTCCACCTGGCTCCAAACTGCTG
GTATTACAGGGTGTGCCACTCTGCTGGCTCATGCTGTGAATGTCAACAGCAATCA

6732 TGAATAGCTGGGATTACAGGCCTGCGCACCATGCCAGCTGATTGGTATTTGGTA
GAGATAGGGTTTCAACCATGTTGGCCAGGCTGGCTTGAACTCTGACCTCAGGTGATCCG
CTGTCTGGCTCCAAAGTGGGATTATAGGCATGCCACTGCAACAACTCCAAA
GCAGCATCTTGTGCTCCCTTCAAGGGCATCACAGAGGGCTGTTGGGTTG
ATGAGAGGCGAAGAATCAGGCATGGAGTCCTCTCAGACTCCCTCTGAGAAGTGG
[G, A]
TGCAGGGGTGGAGAGAAAAGAAGACTAGGCATAGTGGCTCATACCTGTAATCCCAACATT
TTGGGAGGCTGAGGCAGGAAGATTGCTTGAGCTCAGGAGTTGAGACCCAGCCTAGGCAAC
ATAGTGAAGACACATCTCTAAAAAAAGAAAAAAAAGAAAAAAATGAGCCAGGTGAGTGA
CTCATGCCCTGGTCCCACCTCTCCGGAGGCAAAGGTGGAGGATCTTGAGGCTGAG
AAATCGAGGCTACAGTGAGCCATGGTGGCACCACTGCACTCCAGCCTGGAGACAGAGAG

7558 AAGAATCGTTGAACTCGGGAGGCGGAGGTTGCAGTGAGCTGAGAACATGCCACTGCACT
CCAGCCTGGCAACAAGAGCGAAACTCTGCTCAAAGAAAATAATAAAAATAAAAA

FIGURE 3R

AAAAAAAAGGGGGGATATGGGTGAAGTATGGACAAAATAGTGGGCAGGCACAGAT
 GATCTGGACACAGGAGCCCTGGAGTTATTCTGAATCTAAGTGTCACTTTATTAAA
 TATTGTGGCATACACCTACAACAAACATAGCCAACACACCTCCTTTGGAGCTTATC
 [G, A]
 AAGTTCCCCTGTTAAGATTTCCCGCTTGTGATGCCGGTGGGTGGTGTAA
 GCAGGCTTACGGGTGGAGTTCTCACAAAGCATTAACTGCCCTGTCCTAGGTCTGC
 CTTCAGCGAGGATGACGGTACTGCCAGGGCAAGCCTCCATGGGCCAGCTGGCCCTCTA
 CCTGCTCGCTCAGAGCAACTGTGAGTTGTCAAGGGCCACAAGGGGACAGGCTGGT
 CTCACAGCTCAAATGGTCTGGAGGATGAGAAGAGGCCATTGGTGAGCAGACACCATC

 7931 GGTGGCAGTTCTCACAAAGGCATTAAGTGGCCTTGTCCAGGTCTGCCCTCAGCGAGGA
 TGACGGTGACTGCCAGGGCAAGCCTCCATGGGCCAGCTGCCCTACCTGCTCGCT
 CAGAGCCAAGTGTGAGTTGTCAAGGGCCACAAGGGGACAGGCTGGTCTCACAGCTCAA
 ATGGTTCTGGAGGATGAGAAGAGGCCATTGGTGAGCAGACACCATCCGCTGGGGTGG
 GGAGCAGCTGGGGCTCATCAGATGATATTCTCAATGAGAATCAGAACATTGGGTT
 [T, A]
 CTCCCCAGGGCTTCCCACCATCCATTCTGCCATCTCACTGCCCTACGTAGAGGCTCG
 AACCTGTCCCCATAGCCATCTTGACCCAGCTTCCCGCTGCACACATACTATTGAC
 AGGTGTGTTCTGTGGTTTTGTGTTGTTGTTGTTGAGTTGGAGGTTGC
 TCTTGCTGCCAGGCTGGAGTACAATGGCGCAATCTCAGCTACCGCAATCTGCCCTCC
 TGGGTTCAAGCAATTCTCTGTGCTCAGGCCCTGAGTAGCTGGGATTACAGGCATGCC

 8190 ATCAGATGATATTCTCAATGAGAATCAGAACATTGGGTTCTCCCCAGGCCTTTCC
 CACCATCCATTCTGCCCATCTCACTGCCCTACGTAGAGGCTCGAACCTGCCATAGCCA
 TCCCTGACCCAGCTTCCCGCGCTGCACACATACTATTGACAGGTGTGTTCTGCTGG
 TTTGTTTTGTTGTTGTTGTTGAGTTGGAGGTTGCTCTGCTGCCAGGCTGG
 AGTACAATGGCGCAATCTCAGCTACCGCAATCTGCCCTGGGTTCAAGCAATTCTC
 [T, C]
 TGCCCTAGCCTCCTGAGTAGCTGGGATTACAGGCATGCCACACCCAGCTAATT
 GTATTGTTAGTAGACGTGGGTTCTCATGTTGCTCAGCTGGTCTCGAACCTGACC
 TCAGGTGATCGCTGCCCTAGCCTCGAAAGTGTGGGATTACAGGCATGAGCCACTGC
 CTTAGGCCACTGACAAGCCTGTATTGGCTAGCCACCAAGATTGACTGATTATCCACC
 TTCGGGACAACGGACAGCCTGCTTATGACTTACGCCATAGTCTGTCTACTAGCTCTC

 8652 TACAGGCATGAGCCACTGCGTTAGGCCACTGACAAGCCTGTTGGCTAGCCACCAAG
 ATTGACTTGATTATCACCTTCCGGACAACACTGGACAGCCTGTTATGACTTACGCCATAG
 TCTGCTCTACTAGCTCTGCCCTGACTTGACCCAGCATAACAGCCAGAGGCCAGCC
 TTTTCAATATAAACCTGATCTTGCTGCCACTGCTAAACCTGCAAGGGGCTCGCACTGC
 TCCATGGCCCAAGCCTGCTACCCCTACCTCTGCCAGGCTCTGCTCATCCATTCTCTGC
 [C, T]
 TCCCACACACCTGCCCTGTGGCTCCAGGCATACCATCTCAACTCATAAGCCAGTT
 TTTTCAACAGGCCCTCCATCTGGACTGGCTTCCCTGCGTGCAGTTCACTCTGCTCT
 ACCTTGCGACTGCCACCCATCCTCAGCGTCCAGCATTACCTCTGGAGAACATC
 CTGCGACTGCCACCCAAATACACTACTGCTGCTGCAATTCTCGTTGCAATTG
 CAGTCGACATGAGCAATTGCTGTGGTTGAGGCCAAGTGCCTGCTGCCATG

 9370 AGGCCAGGGTCCCAGGTCTGGCGGGCTGGCTGCTGGTGGGGCAGAGAGGCCACCC
 TCTGTTTTTCCCTCTAGGGCATGATCACAAAGGGCACCCCCACACTAGCTACTACCA
 GTATGGCTGGCATTCTGCCCTGTGCTCCACAGAAGCCGGTCCATGACAGCGTGGT
 GGACAAACTCTGTATGCTGGAACCTTCCACCCAGGGCACCATTGTGGGTGAGTA
 GGTCAGACCGTCCAAGGCCAGGCTGGCACTCCCTCAGTCCAGGTCTGCACTGATGAC
 [G, C]
 TCCATACCCCTGGCCCCACACTCACCTTCTGGGCTCTCCGAATCAAGTCCTTAG
 GGACGAATTGGCGAGGGCTCATGGGTGATGCTCCAGCTGAGCCAGCTGGAGCTGGT
 AGGTGGATCTTGAGGCCAGGAGTTCAAGACAACGTGGTAAACCCCATCTACTAAA
 ATAAGGAGGTTAGCGGGCATGGTGGCACATGCCCTGAGTCCCAGCTACTGGGAGGCT
 GAGGAGGAGAATCACTTGAACCTGGGAGGCCAGGTGAGTGAGGAGATCGCACCA

 9463 GGGCCACCCCCACACTAGCTACTACCAAGTATGGCCTGGCATTCTGCCCTGTCTCCA
 CCAGAAGCGGGTCCATGACAGCGTGGTGACAAACTCTGTATGCTGTTGAAACCTTCCA
 CCAGGGCCACCATCTGTGGTGAGTAGGTCAAGCCAGGCTGCCAGGCCAGGCTGGCACTCC
 CTCAGTCCCCAGGTCTGCACTGATGACGCCATACCCCTGGCCCCACACTCACCTTCC
 TGGGGCTCCCGAATCAAGTCCTTACGGACGAATTGGCAGGGCTCATGGTGATGCT
 [C, T]
 CAGCTGTGAGGCCAGTTGGAGCTGGTAGGTGGATCTCTTGAGGCCAGGAGTTCAAGACA
 ACGTGGTGAACCCATCTACTAAAGAAATAAAAAGTTAGCCGGCATGGTGGCACATG
 CCTGCTAGTCCCAGCTACTCGGGAGGCTGAGGCCAGGAGAATCACTTGAACTGGGAGGCG
 AGGCTGAGTGGAGATCGCACCACTGCCCTCCAGCCTGGCAACAGAGTGGAGTGAG
 ACTCTGTCTCAAAAATAAAAAAACTCCCCTAGTGATTCCAATGTGAGCT

FIGURE 3S

9883 GCCTGTAGTCCCAGCTACTCGGGAGGCTGAGGCAGGAGAACACTGAACCTGGGAGGCG
 GAGGCTGCAGTGAGTGGAGATCGCACCACTGCCCTCCAGCCTGGGCAACAGAGTGAGTGA
 GACTCTGTCTAAAAAATAAAAATAAAACTCCCTAGTGATTCCAATGTGCAGC
 TAAGTTGGAAATAGGTGGTATGGGTCAAGTCCTCTGGGCTCCCTCCAGTCCTT
 CTCCCTAACCTCTAGCCCTCAAGTTGCAGAGTGAATCAGCAAACCAGTTGCCAGAAAT
 [G, T]
 AGCAGTTTCTGGGACACAGGATTTCAAGTCCAGACAAGGAAAGTCTTGGGAGACCA
 GGTTGAGTTGGTGCCTTAGCTGATCTGACCATGTTGCCCTCTTCTCCAAGGCCCTCCTG
 TGGTTGTCCATAGCTACAAGGGCCTGACCTCAAGCCCCCTGCTGTCCGGCCCTTTGG
 CTCTCCAGCTCATTGATGTTCTGCCCCACTTCAAGACACAGCAGGATGGCAGGCTT
 GCATTACACCTGTCTGAAGCGCTCAAACCTCAACCCGGAGACAAACGGATCACCAT

 11594 AAAAAAAAAAAATGGAGAAGAAGGAAGCTGGACATGGTGGCTCGTGCTTATAATCCTAGCA
 CTCTGGGAAGCTGAGGAGATGGATTGCTGAGGCCAGAGTCCATCTGAGGCCCTGGGCA
 ACATGGTAAACCCCTGCTTTACTAAATAACGAAAGATTAGCCAGGCATGGTGTAGACA
 CCTATAATCCAGCTACTAGGGAGGCTAGCCACAAGAATCACTTGAACCTGGGAGACAG
 AGGTTGCAGTGAGGCCAGATCGGCCATTGACTCCAGCTGGGAGACTGTGAGACTC
 [G, A, T]
 GTCTCCAGAAAAAACAGAATGGATAGAGTGGAGGCAAGAAGAGGCAGGAAGAACAAAGA
 CACAGAGGTGCACAGAGTTGGGAAATTGGAGGAATGCTTGCACAAAGAGTGGGATC
 TGGGAGAATGAGTGGAGTGGAAAGCAGATGAATGAAGAGAAGGTGAGCGCATCAGGTA
 ACAGAGATGCGTTGTGAACAAATGATGTTCTAGGAAGAGGCCCTGGAGTGTAGGTGC
 CAGAGAGGTGGGAGGAAGGATACTGGAAGCAGAAACAGTGAGGGCCTGATCTTGGG

 12361 ACAAGAAGACAAATAATCCAGGCTCTGTCTCACACCAGCTGCCGCCCTTCTTCC
 TGGCACAGTCATGTTGAAACCAGCTGCTGAGACCATTCTCAGACCCAAGAGATCATCAG
 TGTCACGCTGCAGGTCTTAGTCTCTGCCGCCGTACAGCAGTCATCTGTTCTGGC
 CGGGTCAACCGTGAAGATGCTCTGAAGAACGGCCATGAGTTAGGAGGATTCACTGAGA
 CTCCCCACTCCAGTCTCACCCACCCACCTCACATGCCATGATAACAGGGTCACAGAA
 [A, G]
 AGACGGGAAACAGAGGAGAGGGTTCCCTGGGAGAGACACTGGCCCTGCTCTGCTTCTA
 CCTGCTCAGCTCTTCTTGCCCACGGTTATGAAACAGGGAGCCATAGGCCAGCATT
 GTCACTGAGAGAGCAGGTTGGAGGAGCAGCCCCCAGTGGAAATCCAACCTAACC
 GCTAGGTTCCAGGTAGGCACCCACAATTCAACGAGGAGAACAGTTGTGCCCTTCCCTGC
 AGGGCAGTGTGAAGAGTCCAGGAGTTAGTACACATAGAGATAGTGGCATGTGCTTTTA

 13187 GGCACGTGCCACACAACCCCTGGGTAATTTTTTTTTTTGAGATAGGGCTCTG
 TCTGTTGCCCAGGCTGGTCTCAAATTCTGGCCTCAAACCATCCTCACACCTGAGGGCGCT
 CAAAATATTGGGATTATAGGTGCGAGCCATCATGCTCAGCCAGAAATAAACTGGTTTT
 TTTGTTTTTTTTTGAGACAGAGTCTCACTCTATTACCCAGGCTCTGGAGGCCAACCTCG
 TGTTTGTGTTTTGTTATTATTATTATTATTATTGAGACAGAGCCTCTCTCTTT
 [T, C]
 ACCTAGGCTGGAGTGCAGTGGCGCAATCTGGCTCACTGAAACCTCCGTCTCTGGGTT
 AAGTGATTGCTCTGCTCAGCCTCTGAGTAGCTGGTGACAGGGCGTGCCACCATGC
 CCAGCTAATTGGTATTGGTATTAGTAGAGACAGGGTTTACTATGTTGGCCAGTGGTT
 TAACTCCCTGAACTCGGGTGAATCGCCTGCCCTGGCCTCCAAAGTGTCTGGGATTACAGGC
 ATGGGCTCCGTGCCGCCATGTATTATTAGGCAAGGTCTCTGTATCCAGGC

 13284 ACCATCCTCACACCTGAGGCCTCAAAATATTGGGATTATAGGTGCGAGCCATCATGCTC
 AGCCAGAAATAACTGGTTTTTTGTTTTTGAGACAGAGTCTCACTCTATTAC
 CCAGGCTCTGGAGGCCAACCTGTTGTTGTTGTTTATTGTTTATTATTATT
 TTCGAGACAGAGCCTCTCTTCACTGAGTGGAGTGCAGTGGCGCAATCTGGCTCA
 CTGCAACCTCCGTCTCTGGGTTCAAGTGAATTGCTCTGCCCTGAGTAGCTGG
 [T, C]
 GCTACAGGCGCGTGCCACCATGCCAGCTAATTGGTATTGGTATTAGTAGAGACAGGGTT
 TACTATGTTGGCCAGCTGGTTCTAACTCTGAACCTGGGTGATCTGCCCTGCC
 CCCAAAGTGTGGATTACAGGCATGGCCTCCGTGCCGCCATGTTATTATTAGGCA
 AGGTCTCTCTGTTATCCAGGCTGAAGTGCAGTGGCACATTCAAGCTCACTGCAGCCT
 CAAATTATCCAAGTAACAGGGACTACAGGCATGCACCAACACCACATCTACTTTTT

 21341 TCAGCTCACTGCAACCTCTGCCCTCCCTGGCTTAAGCGATCTCCACCTCAGCCTCTGA
 GTACCGTGTGACCATAGGCCATGGCACAAGGCCAGCTAATTGGTATTGGTATTAGTAGA
 AATGTGGTTTCACTGTTGCAAGGCTGGTCTGAACTTCTGAACGTGATCTGCC
 TGCCCTGGCCTCCAAAGTGTGGGATTCTAGGTATGAGGCCACCCCTGCTGCCCTATAAT
 GGCACCTTCCATTGATGAGGCTACTCTCATGACCTAATCATCTCCAAAGGC
 [C, G]
 CTAAGGCCTCTGATACCATCACCTTGGGTTAGGTTAACATATACATTGGGGG
 ACACAGACATTAGACCATAGCACCTCCATTGAAAGGAAACATTCTGACACCTGGCTA

FIGURE 3T

TCTCAAAGGCCCTTCAGTCCCCGTGAGGCTGCATCCCCACATCCAACAAGAGCAG
 CGACACTCACTCAGAGTTAATAACTTGTCCAGACTCACAGCAGTAATGAATGACAGAG
 CTGGGGCTTGAATCCAGGCGTCCCTAGAGCCTGGATTCTGTGAGTGAAAGCTG

21796 CATTCCCACATCCAACAAGAGCAGCAGACTCACTCAGAGGTTAAAACCTTGTCCAG
 AGTCACAGCAGTAATGAATGACAGAGCTGGGCTTGAATCCAGGCGTCCCTAGAGCCT
 GGATTCTGTGAGTGAAGCTGACTCCCTGGGAGACTCTGCGTGGCTCTGGTTCTC
 TCTCCAGACTCAGTGCAGTTCTCTCCTGATGGTCCCTAGGGTATTACAAAGACA
 GTGGGCCCTGCCTGTCAGGTGTTTATTACAGATGAGGTATGGCCCTCAGGAACCCCTGT
 [A, G]
 GGAAGCTGAGGTCAGAGTCTTGAGCAGGCTTAGGGAGGTTCCAGCTCCCACCACCAA
 GCCCCAGGTGGATTCTTACAGACTCTAGCCTCAGGGGGGCTGGAGATGAGGTTG
 CGGGGTGCGATATTCTGCCAATTGCCCTCCCTGCAATCTGTTCTGAGGTATTG
 CTGACTACAGACCCAAGGATGGAGAACCATTGAGCTGAGGCTGGTAGCTGGTAGCCCC
 TGAGCTCCCTCATCCCAGCAGCCTCGCACACTCCCTAGGCTCTACCCCTCCCTGATG

22159 CCCAGGTGGATTCTTACAGACTCTAGCCTCAGGGTGGGGGCTGGAGATGAGGTTGCG
 GGGTGCATATTCTGCCAATTGCCCTCCCTGCTCAATCTGTTCTGAGGTATTGCT
 GACTACAGACCCAAGGATGGAGAACCATTGAGCTGAGGCTGGTAGCTGGTAGCCCCCTG
 AGCTCCCTCATCCCAGCAGCCTCGCACACTCCCTAGGCTCTACCCCTCCCTGATGTC
 CCTGGAACAGGAACCGCCTGACCCCTGCTGCCACCTCTGTGCACTTTGAGCAATGCC
 [C, T]
 TGGGATCCCCAGCACAGCCCTCGAGGGCCCTACCATGGCCACCTGGAGCAG
 AGAGCCAAGCATCTCCCTGGAGCTTCTGCCAAGTCTGGCAGCTGGCCCTGCA
 GGTCTCCCAGAAGGCCACCCATGGTCTGATGGGATGAAGCAGTCTCAGACTCTGGC
 AAAAAACGGAGTCGCAGGGCGCAGGTGTTGTGAAGACCACTCGTCTGTGGTGGGTC
 CTGCAAGAAGGCCCTCAGCCGGGGCTATGGCCCTGACCCAGCTCTCCACTCTGCT

22334 CCCTGAGCTCCCTCATCCCAGCAGCCTCGCACACTCCCTAGGCTCTACCCCTCCCTG
 ATGTCCTGGAACAGGAACCTGCCTGACCCCTGCTGCCACCTCTGTGACTTTGAGCAAT
 GCCCCCTGGGATCCCCAGCCACAAGCCCTCGAGGGCCCTACCATGGCCACCTTG
 GAGCAGAGAGCCAAGCATCTCCCTGGAGCTTCTGCCAAGTCTGGCAGCCTGGC
 CCTGCAGGTCTCCCATGAAGGCCACCCATGGTCTGATGGGATGAAGCAGTCAACTC
 [C, A]
 TTGGCAAAACGGAGTCGCAGGGCGCAGGTGTTGTGAAGACCACTCGTCTGTGGTTG
 GGGTCTGCAAGAAGGCCCTCGAGCCCCGGGCTATGGCCCTGACCCAGCTCTCCACT
 CTGCTGTTAGAGTGGCAGCTCGAGCTGGTGTGGCACAGTAGCTGGGAGACCTCAGCA
 GGGCTGCTCAGTGCCTCTGACAAAATTAAAGCATTGATGGCCTGTGGACCTGCTAC
 ACTGGCCTGGCCTCATCTCTCAGTGCAGGGCAGGGACAAGAGAAGGGGAAGTAACCC
 ACTCAGGG

22346 TCATCCCAGCAGCCTCGCACACTCCCTAGGCTCTACCCCTCCCTGATGCCCCGGAA
 CAGGAACCTGCCTGACCCCTGCTGCCACCTCTGTGCACTTTGAGCAATGCCCTGGGAT
 CACCCAGGCCACAAGCCCTCGAGGGCCCTACCATGGCCACCTGGAGCAGAGGCC
 AACCATCTCCCTGGAGCTTCTGCCAAGTCTGGCAGCCTGGCCCTGAGGTCTC
 CCATGAAGGCCACCCATGGTCTGATGGGATGAAGCAGTCTCAGACTCCTGGCAAAAA
 [C, -]
 GGAGTCCGCAGGGCGCAGGTGTTGTGAAGACCACTCGTCTGTGGTGGGCTTGCAAG
 AAGGCCCTCTCAGGCCGGGCTATGGCCCTGACCCAGCTCTCAGTCTGTTAGAG
 TGGCAGCTCCAGCTGGTGTGGCACAGTAGCTGGGAGACCTCAGCAGGCTGCTCAGT
 GCCTGCCCTGACAAAATTAAAGCATTGATGGCCTGTGGACCTGCTACAGTGGCCTGGT
 CCTCATACTCCCTCAGGTGCAGGGCAGGGACAAGAGAAGGGGAAGTAACCCACTCAGGG

22476 ACAAGCCCTCGAGGGCCCTACCATGGCCACCTGGAGCAGAGAGCCAAGCAGTCTC
 CCTGGGAAGTCTTCTGCCAAGTCTGCCAGCCTGCCCTGAGCTCTCCATGAAGGC
 CACCCATGGTCTGATGGGATGAAGCAGTCTCAGACTCCTGGAAAAAACGGAGTCCGC
 AGGCCGCAGGTGTTGTGAAGACCACTCGTCTGTGGTGGGCTCTGCAAGAAGGCCCTCC
 TCAGCCCCGGGCTATGGCCCTGACCCAGCTCTCAGTCTGCTGGTAGGTGGCAGCTC
 [C, T]
 GAGCTGGTTGTGGCACAGTAGCTGGGAGACCTCAGCAGGGCTGCTCAGTGCCTGCC
 GACAAAATTAAAGCATTGATGGCCTGTGGACCTGCTACAGTGGCCTGGTGCCTCATA
 CTCAGGTGCAGGGCAGGGACAAGAGAAGGGGAAGTAACCCATCAGGGAGGAGTGGAG
 GGTGCCTGAGCCGCATGTGGCATTGGGGAGTGTGGGAATGCCAGCAGTGTGACGT
 TGACTACTGACTGAGCACCCACTACTATGACTGAGCAGTCACTCGCTAGATACTATCTT

23143 GCGGGCGTGGTGGCTACGCCGTAACTCCAGCAGTCTAGGAGGCCAAGGCAGGTGGAT
 CAAAGGTAGGAGTTTGAGATCAGCCTGCCAACATGGTAAACTCCATCTTACTAAA
 AATACAAAAAATTAGCCAGGCATGGTGTGATGCCCTGCTGAGTGGCAGCTGAGTGGCAG
 TGGGAAGCTGAGGCAGGAGATTGCTGAAACCTGGAGGCGAGGGTTGTAGTGAGGCC
 ATCACGCCATTGCACTCCAGCTGGCAAGAAGAGAACACTCTCAAAAAAAAAAAAAA

FIGURE 3U

[A, -, T]
CAGGAAACTGGTGCCTAAAAAGGAAAAGTGAACCAAGGTACAGACTAGGCAGTGAT
GCTGGGGGAACCTGGCTCAGGGCACAGACCTGGCTGGGCAGCCTTGCAGCTCCCTCC
ACTAAAATACTGAAAATGAGGGGCTTCGATGATGGTTATAATCGTATGGCAGAGCCCCAA
CTCAACTGGAGCCCTGGGACCCAGAAGCTAGGGTCTCACTCCCTGCTTTCCACAAGGCA
CCATTAGGGCATACCCCCAGGCCCTCGCAGCCACGACGAGGGATCCTGCCTCTCATTGG

23445 AGGAAACTGGTGCCTAAAAGGAAAAGTGAACCAAGGTACAGACTAGGCAGTGATG
CTGGGGGAACCTGGCTCAGGGCACAGACCTGGCTGGGCAGCCTTGCAGCTCCCTCCA
CTAAAATACTGAAAATGAGGGGCTTCGATGATGGTTATAATCGTATGGCAGAGCCCCAAAC
TCAACTGGAGCCCTGGGACCCAGAAGCTAGGGTCTCACTCCCTGCTTTCCACAAGGCAAC
CATTAGGGCATACCCCCAGGCCCTCGCAGCCACGACGAGGGATCCTGCCTCTCATTGGT
[T, C]
GGGGGCTTAGGGCTCTGGCTGCCCTTGAAGAGGGGTTCAAGCCCAGCGAGGCACCC
CCTATGCTGCACCCCCACCAAGGTTAGGAAGAGGTCTGCTCAGTGGGGCCCTCTGATG
AACAGCCCCATCAGGCTCGCTCCACATGCCCTGGAAGAGATGGTACATACTCAAAGTCC
TTGAAGCCGCATAATTAAACACCTAGAGCACCATCTAACATTAGGGTCTGAGAAGA
TAGGGGAAGTAAGCAATTAAAACATTCTTATATTGGCCAGGTGCAATGGCTCACGT

23974 GGTCTGAGAAGATAGGGAAAGTAAGCAATTAAAACATTCTTATATTGGCCAGGTGC
AATGGCTCAGCTCTGAATCCCAGCCCTTGGGAGGAGGATCACCTGAGGTCAAGGAG
TTCAAGATCAGCTGGCCACATGGAGAACCCCCATCTACTAAAAATACAAAAAATTAG
CTCAGGCCTGTTGATGTGACCTGTAATCTTAGCTATTCAAGGAGGCTGAGGCAAGAAAT
TGCTTGAGTCAATATTGACCACTGCACCTCCAGCCTGGCAACAGCAGACTCTGTCTC
[A, -]
AAAAAAAAAAAGATATTGCTGAAAAGACCCAGCCTGCCAAACTCAGGGCAGCCAAGG
GAGGTAGTGAATGGAAGTGGAGGCTCAGCCTGCCACACCTCCACTGCCCTCAGGCCCT
CTCTGCCTCTTCCCACAGTCTGCTCTGGCATCTGCTCTGGCAGAGACTTGGCCT
CTTCCAGTTCAGCTCCCTCTTAGATTGTGTCCTGCCACTGAGTCTTGGGACACTG
GGTCAGATGCTAGTCTGGCACAATTGGCAGGAATCCAAGAACAGTGTGAGTGGGG

24865 ATAGCTGGGATTACAGGTGTGCACCAACCATGCCAGCCTAATTTTGATTATTAGTAGA
GATGGGGTTTCACCATGTTGTCAGGCTGGTCACTGACTCCTGACCTCAAGTGTACCC
CGCTTGGCCTCCAAAGTGTGGATTACAAGCATGCCACAGTGCCTGGCCTGACCC
TCTCTTTGAAAGCCTCCAAATCTGTGCACTGTGTGCCCTTCTCTCTG
CCTCTCTCAGCTCTGCCCGCTCTCCCTCTCTGGCAAATCCACTCATCTCT
[G, T]
GAAGCCCTTCCAGGGAAAGCCCTGATCATGCTGTTCTCCCTGTGGAGGGATGAAG
GACGTGGCCCACGGAGTTGTTGTTGTTGAGATGGAGTTTGCTCATGTTGCC
AGGCTGGGTACAATGGTACGATCTCAGCTCACTGCAACCTCTACGTCCGGTCAAGC
GGTCTCCTGCCCTAGCCTCCCAAGTAGCTGGGATTACTGGCATGAACCACACCTGG
CTAATTGTTGTTTGTAGAGATGGGTTCTTCATGTTGGTCAGGCTGGTCTCGAAC

25040 GACCTGCTCTTTGAAAGACCATCCCCAAATTCTGTGCACCTGTGTGCCCTTCTTCT
CTCTGCCTCTCTCAGCTCTGCCCTCTCCCTCTGGCAAATCCACTCA
TCTCTGAAGCCCTTCTCAGGGAAAGCCCTGATCATGCTGCTTCTCTGTGGAGGG
ATGAAGGACGTGGCCCACGGAGTTGTTGTTGAGATGGAGTTTGCTCATGTT
TTGCCAGGCTGGGTACAATGGTACGATCTCAGCTCACTGCAACCTCTACGTCCGGGGT
[T, C]
CAAGCGGTTCTCCTGCCCTAGCCTCCCCAGTAGCTGGGATTACTGGCATGAACCAACCA
CCTGGCTAATTGTTGTTTGTAGAGATGGGTTCTCATGTTGGTCAGGCTGGTCT
CGAACTCCAAACCTCAGGTGATCTGCTGCCCTGCCCTCCAAAGTACTGGGATTACAGG
GTTGAGGCCACTGTGCTGCCAGGGCCACGGAGTTAAGAGGCTCTGTGGCAGTGG
CATCCAGACGGAGTGCAGAAACTCAAAGTTGAAGGCCAGAAGCTCAGGGAGGGGGAGTG

25046 GCTCTTTGAAAGACCATCCCCAAATTCTGTGCACCTGTGTGCCCTTCTCTCTGC
CTCTCTCAGCTCTGCCCTCTCCCTCTGGCAAATCCACTCATCTCT
GAAGCCCTTCTCAGGGAAAGCCCTGATCATGCTGCTTCTCTGTGGAGGGATGAAG
GACGTGGCCCACGGAGTTGTTGTTGTTGAGATGGAGTTTGCTCATGTTGCC
AGGCTGGGTACAATGGTACGATCTCAGCTCACTGCAACCTCTACGTCCGGTCAAGC
[G, A]
GTTCTCCTGCCCTAGCCTCCCCAGTAGCTGGGATTACTGGCATGAACCAACACCTGGC
TAATTGTTGTTTGTAGAGATGGGTTCTCATGTTGGTCAGGCTGGTCTCGAAGT
CCCAACCTCAGGTGATCTGCCCTGCCCTGCCCTCCAAAGTACTGGGATTACAGGGTTGAG
CCACTGTGCCCTGCCAGGGCCACGGAGTTAAGAGGCTCTGTGGCAGTGGCATCCA
GACGGAGTGCAGAAACTCAAAGTTGAAGGCCAGAAGCTCAGGGAGGGAGTGAGTT

25193 ATCATGCTCTTCTCTGTGGGAGGGATGAAGGACGTGGCCACGGAGTTGTTGTTGTT
TTGTTTGAGATGGAGTTTGCTCATGTTGCCAGGCTGGGTACAATGGTACGATCTCA

FIGURE 3V

GCTCACTGCAACCTTACGTCCGGGTTCAAGCGGTTCTCCTGCCTTAGCCTCCCCAGTA
 GCTGGGATTACTGGCATGAACCACACACCTGGCTAATTTGTGTTTTAGTAGAGATGG
 CGTTTCTTCATGGTCAAGGCTGGCTCGAACCTCAACCTCAGGTGATCTGCCCTGCCT
 [C, T]
 GGCCCTCCAAAGTACTGGGATTACAGGGTTGAGCCACTGTGCCTGCCAGGCCACGGA
 GTTTTAAGGGCTCTGGCATCCAGACGGAGTGCAGAAACTCAAAGGTTGAA
 GCCCAGAAGCTCAGGGAGGGGAGTGTGAGTTGAGGAGTCTTGGCTGCCAGGCCAG
 AAACCGAACCTCAAGCCTCTCCACAACAGCGGGTGTAGAGCATGTAGAATCAGAGAGGAG
 GCTGAGCCATGCAGCCCCAGAGAGGGAAATGCCACTGAGCCACAGAGACCCAGTGCCA

25529 AGTGCAGAAACTCAAAGTTGAAGGCCAGAAGGCTCAGGAAAGGGGAGTGTGAGTTGAGGA
 GTCTCTGGCTGCCAGGCCAGAAACCGAACCTCAAGCCTCTCCACAACAGCGGGTAG
 AGCATGTAGAATCAGAGAGGAGCTGAGCCATGCAGCCCGAGAAAGAGGGAAATGCCACT
 GAGCCACAGAGACCCAGTGCCACTGCCAGGTGTCTCTGCCACTTCCATGACCC
 [T, G]
 GCTGTCTCTGTATGCCAGGTTCACCCCTCTCGTTGACATTGTACACATTCTAGGTGA
 CACCAGCAGCTCTGATTCTCATCTCCATAACATCAGCCCCCAGAGAGGGAACTG
 CTGAGCTGATAACATAATAGATGCCCTTCTGGAGGCCATGGTCATGGTCAGCGTGG
 GAGGATGAAGCCTGAGCAGGCAGGATCGGGGTCTAGAGGGAGGGTGAAGGTT
 25614 GCCCAGAAGCTCAGGGAGGGGACTGTGAGTTGAGGAGTCTTGGCTGCCAGGCCAG
 AAACCGAACCTCAAGCCTCTCCACAACAGCGGGTGTAGAGCATGTAGAATCAGAGAGGAG
 GCTGAGCCATGCAGCCCGAGAAGAGGGAAATGCCACTGAGCCACAGAGACCCAGTGCCA
 CTGCCAGGTGTCTGCCCTCACTTCCATGCCGGCTGTCTGTATGCAGGCTTC
 CCTCTCTCGTTGACATTGTACACATTCTAGGTGACACCAGCAGCTCTGATTCTCATC
 [C, T]
 CCCATAACATCAGCCCCCAGAGAGGGACAACGTGAGCTGATAAACATAATAGATGCC
 CCTTCTGGAGGCCATGGTCATGGTCAGCGTGGAGAGGATGAAGCCTGAGCAGGCAGGA
 TCGGGGTCTAGAGGGAGGGAGGTGGAGATCACAGACCTGTGGTCAGGTGG
 TGGGAAGGGTTTGACGAGTGTGCGCCAAAGAGCTTGAAGGGATTTCGTCAGGTGG
 GAGCACTGCCCTCTCCCTAGGGACAACAGCCACCTCTCTCCCCATTGCTTTCCC

25785 CCAGTGCCACTGCCAGGTGTCTCTGCCCTCACTTCCATGCCGGCTGTCTGTATG
 CAGGCTTCACCCCTCTCGTTGACATTGTACACATTCTAGGTGACACCAGCAGCTTCTG
 ATTCTCATCTCCATAACATCAGCCCCCAGAGAGGGACAACGTGAGCTGATAAACAT
 ATAGATGCCCTTCTGGAGGCCATGGTCATGGTCAGCGTGGAGAGGATGAAGCCTGA
 GCAGGCAGGATCGGGGTCTAGAGGGAGGGAGGTGGAGATCACAGACCTGTGG
 [C, T]
 CAGGTGGCTGGGAAGGGTTTGACGAGTGTGGCCAAAAGAGCTTGAAGGGATTTCG
 GCTGTGGGTGAGCACTGCCCTCCCTAGGGACAACAGCCACCTCTCTCCCCATT
 GCCTTCCCTCTGTAGATATGAAACACAGGCCCTCTGTGAGGCCCTACTAACCTCC
 GTATGGGAAAGCGCCGGAGAAAGGGAGTTCTGGCAGCTCTCCGAGACCCAAACACC
 CCACTGTTGCAAGGTGAGTCATGCCCTGACACTCTGGATGTGCTCTGG
 26266 GTGATGGGAAAGCGCCGGAGAAAGGGAGTTCTGGCAGCTCTCGAGACCCAAACACC
 CCACTGTTGCAAGGTGAGTCATGCCCTGACACTCTGGATGTGCTCTGG
 CTCAGCCAAGAGGCTTCATCAACTCACCCAGCTTCCCTAGCACCCCTCTGG
 CACACTTCACAAATCACTGATGCTCAAAGTTGGATATAATATATTGAACTGAAGCCT
 TAGCATTTATGCAAGTTACTGTGAAATTCTAGGAAACCAGACAGATTACAAAA
 [A, -]
 CTAGAAGAAAATTAACATCACCTAGGATATACTACCTAGGAATAACGTCTTTATTTGA
 GATGGAGTTCGCTTGTGCCCAGGCTGGAGTGCAGCGTATGATCTCGCTCG
 AACCTCCGCCCTCTGCCCTCCAAAGTTCTGGATTACAGGCATGAGCCACCG
 CACAGGCCAGTGGCTGCTCTGG
 AAAATTACTTAACCTTCTCTAGATACTTTTAAATATGGCAGTAAGTTTCATAA

FIGURE 3W